

WORKING PAPER

Meters on the Roof

A Statistical Examination
of the Project Results

CONSUMER AND
CORPORATE AFFAIRS CANADA
LEGAL METROLOGY BRANCH

Canada

H-MON-1984157, E, C

Legal Metrology

MEMORANDUM

NOTE DE SERVICE

TO
A

Corinne MacLaurin
Bibliothécaire en chef/
Chief Librarian

FROM
DE

Directeur
Direction de la Métrologie légale

Director
Legal Metrology Branch

SUBJECT
OBJET

COMPTEURS SUR LE TOIT

Veuillez trouver ci-joint une copie des versions française et anglaise du document intitulé Compteurs sur le toit préparé par MM. M. Romanowski et E. Green pour le compte de la Direction de la Métrologie légale de Consommation et Corporations Canada.

Ce "document de travail" présente une analyse statistique des résultats d'une expérience qui s'est échelonnée sur plus d'une dizaine d'années et qui a pris fin il y a bientôt cinq ans. Ses conclusions demeurent cependant très actuelles et sont susceptibles d'intéresser particulièrement les statisticiens.

Je me permets de conseiller aux lecteurs qui s'intéressent aux analyses de variances de lire la Note à la page 24 du document.

Dans le cas où vous désireriez d'autres copies de ce document, vous êtes priés d'en faire la demande auprès de Hélène Pagé au numéro de téléphone (613) 990-8617.

SECURITY - CLASSIFICATION - DE SÉCURITÉ

OUR FILE/NOTRE RÉFÉRENCE

G6520-5

YOUR FILE/VOTRE RÉFÉRENCE

DATE

84-10-19

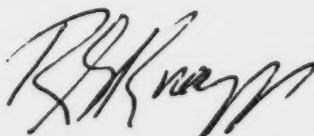
METERS ON THE ROOF

Please find enclosed copies of the English and French versions of the publication entitled Meters on the Roof prepared by M. Romanowski and E. Green for the Legal Metrology Branch of Consumer and Corporate Affairs Canada.

This "working paper" contains statistical analysis of variance on data from an experiment with electricity meters that covered more than a decade and was stopped almost five years ago. Nevertheless, its conclusions will be of interest, particularly to statisticians.

Those readers who are interested in analysis of variance should read the Note on page 22.

If additional copies of this document are required, please contact Hélène Pagé at (613) 990-8617.



Richard G. Knapp

p.j.

encl.

CCA TK301
R6

METERS ON THE ROOF

A Statistical Examination of the Project Results

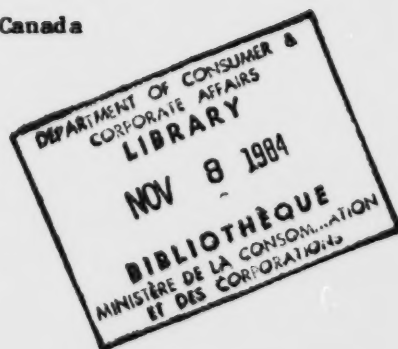
M. Romanowski

and

E. Green

Legal Metrology Branch

Consumer and Corporate Affairs Canada



The analysis and conclusions of this study do not necessarily reflect
those of the department

© Minister of Supply and Services Canada 1984

Cat. No. RG 23-80/1984E

ISBN 0-662-13336-6

Version française disponible

ACKNOWLEDGEMENTS

The authors express their gratitude to R.G. Knapp, Director, for his continual support and encouragement during the various phases of the preparation of this memorandum.

They also thank Lexica Inc. for its help in the typing of the English and French versions of the manuscript.

TABLE OF CONTENTS

	<u>Page</u>
Introduction.....	1
Outline of the Project.....	5
Outline of the Analysis of Variance.....	9
Numerical Calculations.....	15
Other Statistical Investigations.....	19
Appendix I	
Tables I-B to VIII.....	25
Appendix II	
Disk Errors in the Project.....	37

INTRODUCTION

To understand the reasons why the experimental project analyzed in this report was undertaken, it is useful to highlight the events that led to the project and to review the Legal Metrology Branch's policies during the early 1960's regarding meter verification, sealing, and reverification.

The Electricity Inspection Act specified that, before being put into service, all measuring devices used for revenue purposes be verified for accuracy against regulated limits of error and then sealed by a government inspector to prevent tampering. The common household meter is such a device.

Because no meter can be expected to remain accurate indefinitely, the Act also specified that each meter be "reverified" at prescribed intervals. A verification seal period of six years was originally prescribed by regulation for all electricity meters, but this was subsequently extended to eight years for domestic meters. Consequently, until the early 1960's it was common practice for utilities to remove from service all those meters that had reached the end of the seal period, routinely inspect and service them in a meter shop and then have a government inspector reverify and reseal them.

Technological developments in the design and construction of meters -- particularly the advent of magnetic suspension meters -- led both utilities and manufacturers to demand that seal periods be extended. The rationale for this request was that better quality meters would maintain accuracy within prescribed tolerances over longer periods of time.

During the same period, the Legal Metrology Branch was in the process of introducing and refining statistical sampling plans for meter verification and reverification. Several staff members decided to carry out a number of long-

range experiments that would test the validity of the verification solutions being adopted and also help establish Branch policies regarding the metrology aspects of domestic watt hour meters. In addition, the generalized use of magnetic suspension meters was encouraged by the notion that an appropriately constructed meter could be placed either indoors or outside with no significant difference in performance over time. Branch staff felt that this concept, however reasonable and attractive, should be tested experimentally in a natural environment over a number of consecutive years. Thus was born the "Meters on the Roof" project of 1968.

A total of 144 meters, 36 meters from each of the then four manufacturers were installed on specially constructed racks on the roof of the Legal Metrology Branch Building. Each meter was calibrated and adjusted as near as was practicable to zero error at 2.5 and 25 amperes current before installation for the test program. In 1970, an analysis of the first full year's quarterly observations indicated that the findings were inconclusive because each of the three tiers of meters had been energized from a separate phase of a three-phase system. It was therefore decided to test all meter in series and on single phase. Three calibrated "monitor" meters located inside the building were also placed on the same phase. These meters registered what was deemed to be, for statistical calculations, the "true" number of watt hours consumed by the test meters on the roof.

Following this modification, observations taken quarterly during 1971 were analyzed and described in a Technical Memorandum published in 1975. Quarterly recordings of registration errors continued uninterrupted until October 1979, when the project was terminated. Direct meter readings taken during the months of May and October over the eight-year period (a total sample of 2252) were later converted to the more convenient percentage form.

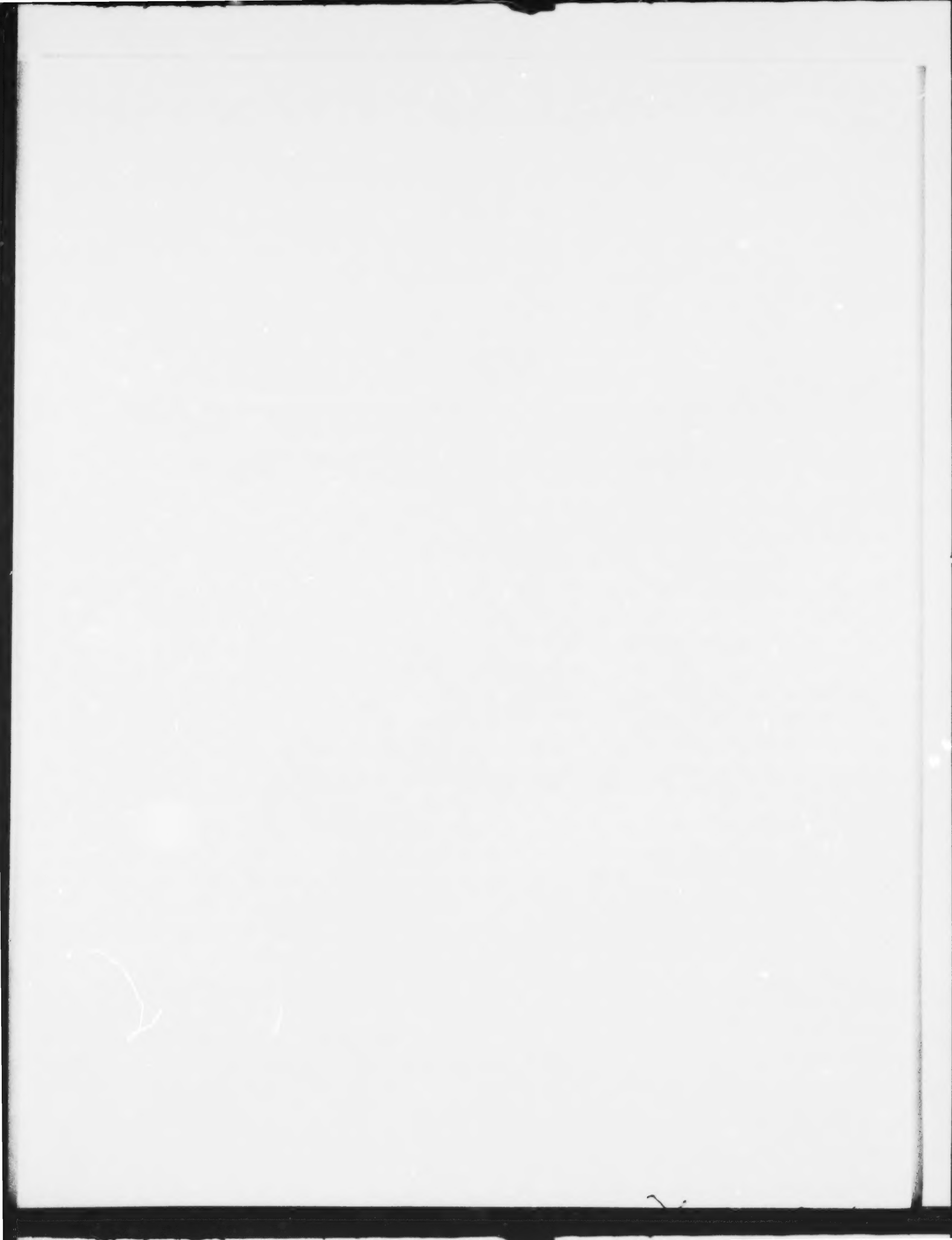
In 1982 the Branch decided to undertake a statistical analysis of variance with respect to the criteria "manufacturer" and "orientation".

The reader, upon examining the results of this analysis of variance, will note the amount and variety of information that the authors have been able to extract from the experimental data. It was possible to break down the 2252 observations into clear-cut subsamples according to precise criteria. It was possible to consider, for instance, observations made on all meters from a particular manufacturer, or observations made on all meters facing north, or observations made during warm seasons. Consequently, the sample was well suited to both an analysis of variance and to a study of the "theory of mixtures" -- one of the major fields of research of the authors of this report (see Note, p. 22).

In conclusion, it can be considered a lucky coincidence that while the analysis of variance was being adapted to the "Meters on the Roof" project, the theory of mixtures was studied and developed, totally independently, at the National Research Council's Division of Physics. The final juncture of these two domains of statistics, and the resulting success, demonstrates the benefit that two institutions can gain through cooperation and mutual assistance.

The Legal Metrology Branch is indebted to Mr. E. Green of the Division of Physics at NRC, and in particular to Mr. M. Romanowski, consultant to the Branch.

Richard G. Knapp
Director
Legal Metrology



1.

OUTLINE OF THE PROJECT

A brief history describing the original objectives and experimental arrangements is contained in the "Introduction" prepared by Richard G. Knapp, Director, Legal Metrology Branch.

From a purely statistical point of view, each year of the eight-year long experiment is represented by four samples of TRE (Total Registration Errors); each sample is constituted by the recordings performed on the totality of 144 meters under test. Each of four manufacturers (General Electric, Ferranti-Packard, Westinghouse, Sangamo) is represented by a group of 36 nominally identical meters, each group being divided into four equal "cells". A cell (9 meters) is thus the smallest set of meters where any two meters can be interchanged without affecting the analysis.

The recordings were performed four times a year: in February, May, August and November. If none of the meters had been eliminated, this would have led to a total of 4608 TRE. In reality, this population was somewhat reduced as the initial number of 144 decreased gradually in the course of the years to 136.

Each "recording" operation, i.e., the reading of all meters represents for the team of two of the Electrical Laboratory staff members 2 or 3 days of work. The reduction of these recordings, i.e., the expressing of them as "percentage errors" is time-consuming work. Because of the shortage of manpower (technical and clerical) initially only the recordings made in November could be reduced. The analysis of variance was thus applied to the yearly results of eight consecutive years (November to November), and the number of the year, e.g., 1972, indicates the TRE (in % form) related to the period extending from November 1971 to November 1972., Table IA.

TABLE IA (1972)

	GE i = 1	FP i = 2	W i = 3	S i = 4	h = 4
SOUTH	$Y_{11} = -0.01$	$Y_{2,1} = +0.72$	$Y_{3,1} = +0.62$	$Y_{4,1} = +1.24$	
	- 0.20	+ 0.56	+ 0.60	+ 1.40	
	+ 0.12	- 0.95	+ 0.09	+ 0.72	
	+ 0.39	+ 0.83	+ 0.76	+ 1.15	
	+ 0.07	+ 0.24	+ 0.14	+ 0.80	
	- 0.17	- 0.40	- 0.72	+ 0.62	
	+ 0.16	+ 0.44	+ 0.73	+ 0.64	
	- 0.01	- 0.91	+ 1.06	+ 1.02	
EAST	$Y_{19} = +0.54$	- 0.49	+ 0.87	+ 0.95	
	$Y_{1,10} = +0.31$	$Y_{2,10} = -0.13$	$Y_{3,10} = +0.05$	$Y_{4,10} = +0.70$	
	+ 0.30	+ 0.27	- 0.29	+ 0.64	
	+ 0.11	+ 0.43	- 0.21	+ 1.20	
	+ 0.65	+ 0.34	- 0.38	+ 0.57	
	+ 0.36	+ 0.07	+ 1.17	+ 0.89	
	+ 0.58	- 0.07	+ 0.60	+ 0.75	
	+ 0.56	+ 0.40	- 0.14	+ 0.92	
NORTH	+ 0.65	+ 0.53	- 0.18	+ 0.87	
	$Y_{1,18} = +0.47$	+ 0.05	+ 0.24	+ 0.81	
	$Y_{1,19} = +0.32$	$Y_{2,19} = +0.26$	$Y_{3,19} = +0.39$	$Y_{4,19} = +0.51$	
	+ 0.11	+ 0.69	- 0.07	+ 0.60	
	+ 0.07	+ 0.98	+ 0.68	+ 0.98	
	+ 0.65	+ 0.55	- 0.18	+ 0.44	
	+ 0.72	+ 0.83	+ 0.59	+ 0.72	
	+ 0.69	+ 0.30	- 0.03	+ 0.25	
WEST	+ 0.68	+ 0.62	+ 0.98	+ 0.92	
	+ 0.62	+ 0.31	+ 0.31	+ 0.81	
	+ 0.35	+ 0.42	+ 0.03	+ 1.20	
	$Y_{1,27} = +0.22$	$Y_{2,27} = +0.50$	$Y_{3,27} = +0.27$	$Y_{4,27} = +0.27$	
	+ 0.46	+ 0.66	+ 0.40	+ 2.50	
	+ 0.36	+ 0.30	+ 0.03	+ 0.59	
	+ 0.05	+ 0.77	+ 0.28	+ 0.45	
	+ 0.42	- 0.03	+ 0.15	+ 0.47	
	+ 0.99	+ 0.50	- 0.43	+ 0.56	
	+ 0.40	+ 0.04	+ 0.56	+ 1.65	
	+ 0.40	+ 0.20	+ 0.57	+ 0.88	
	$Y_{1,36} = +0.63$	$Y_{2,36} = +0.32$	$Y_{3,36} = -0.08$	$Y_{4,36} = +0.45$	
	$t_1 = +13.02$	$t_2 = +10.20$	$t_3 = +9.46$	$t_4 = +30.14$	$t_0 = 62.82$
	$t_1^2 = 169.52$	$t_2^2 = 104.04$	$t_3^2 = 89.49$	$t_4^2 = 908.42$	
					$t_0^2 = 3946.35$

After some calculations had been carried out, it became obvious that for statistical investigations other than the analysis of variance, larger samples would be more appropriate. The decision was then taken to reduce also the recordings of the observations made in May. The calculations presented in Section 4 are performed on samples that, by a general consensus, can be termed "large": their sizes are of the order of two thousand elements. The conclusions drawn from samples of such sizes are considered as sufficiently reliable to lead to practical conclusions. Before applying the methods of statistical analysis let us first examine the nature of several general questions which may be asked and to which the analysis that follows may propose more or less complete answers.

a) General consistency and reliability of the instruments when they are used under special conditions and, by extension, the likelihood that they will behave similarly when placed in conditions in which they are used in practice. It is obviously of considerable importance, for the manufacturer, for the official metrological institutions and, of course, for the consumer of electrical energy to determine what kind of atmospheric conditions influence the functioning of a meter. Is it possible to foresee whether the location of a meter on a house or factory can influence the amount of money a customer has to pay? Does the meter located on a sunny side of a house favour the producer or the user of electrical power?

b) Influence of the criterion of classification "manufacturer".*¹
Although all meters are assumed to be "identically constructed" it is obvious that each manufacturer is likely to introduce some small specific factors the presence of which may not be directly perceptible but may be nevertheless evidenced statistically.

*¹ In statistics, terms such as "manufacturer", "direction", etc., designate "criteria" according to which a sample can be classified.

- c) Influence of the criterion "geographical direction". This is actually one of the major reasons why the experiment has been undertaken. It is impossible to assert a priori that a meter facing South will register electricity in the same manner and with the same accuracy as a meter facing North. Orientation may act not only through temperature but also through such factors as wind, humidity, etc.
- d) Influence of the criterion "season". This poses problems which are, from the standpoint of statistical calculations, identical to those of "direction".

The information supplied through the analysis of these problems is equally important by its implications of scientific, technical and economic nature.

The mathematical method used to compare the behaviour of the meters by taking into account a certain criterion of classification is that of "analysis of variance". The principles of this powerful statistical tool have been established by R.A. Fisher and are, in general, presented in all text books on statistics.

For the convenience of the Reader, however, the fundamentals of the analysis of variance, as they are used in this report, are briefly outlined in the following section.

2.

OUTLINE OF THE ANALYSIS OF VARIANCE

In this report (as in all previous works on similar matters) the symbol Y (capital letter with appropriate subscripts) designates the so-called "Total Registration Error" (TRE) calculated on the yearly basis, i.e., from November to November. A number such as Y_k ranging from 1 to 144 is the "percentage" by which the registration of the k^{th} meter differs from that of the monitors.*² If, for instance, $Y_{10} = -0.23$, this means that the meter of rank 10 registers (after twelve months of being energized) an amount of electricity which is 0.23% smaller than that registered by the monitors*. The latter is sometimes, but not very correctly called "true" registration, in kWh.

To make this report more easily accessible to non-statisticians, only the analysis of variance with one single criterion will be used: the 144 meters will be divided into four classes either according to the criterion "manufacturer" or according to the criterion "direction". At the beginning of the experiment (1972) all classes were exactly of the same size (36 meters). As it has been mentioned above, the initial sizes were not rigorously maintained because in the course of the years some meters were, for various reasons, eliminated but not replaced.

*² Monitors are the reference standards located in the laboratory (not outdoors). They form the base for the evaluation of the registration of the meters under test.

If a sample of N elements Y_k ($k = 1, 2, \dots, N$) is considered in its unclassified form, the mean is given by the formula

$$\bar{Y} = \frac{1}{N} \sum_{k=1}^N Y_k. \quad \dots(1)$$

An important and basic relation in all the calculations that follow is the expression of the sum of squares of the deviations of sample elements from any number A in terms of the sum of squares of the deviations from the sample mean $\bar{Y}$:

$$\sum_k (Y_k - A)^2 = \sum_k (Y_k - \bar{Y})^2 + N(\bar{Y} - A)^2. \quad \dots(2a)$$

For $A = 0$, this reduces to

$$\sum_k Y_k^2 = \sum_k (Y_k - \bar{Y})^2 + N\bar{Y}^2. \quad \dots(2b)$$

If the sample is distributed into "classes", these classes will be denoted by the subscript i , the total number of elements in the i^{th} class will be denoted by the symbol n_i . The rank in the class will be indicated by the subscript j so that $j = 1, 2, 3, \dots, n_i$. An element will therefore have two subscripts, e.g., Y_{ij} :

$$Y_{ij} \text{ with: } i = 1, 2, \dots, h \text{ (i.e. } h \text{ classes; here } h = 4\text{)} \\ j = 1, 2, \dots, n_i \text{ (} n_i \text{ elements in the } i^{\text{th}} \text{ class)}$$

The i^{th} class constitutes in itself an unclassified sample of n_i elements so that it is possible to apply to it the relation (1) in which a) the subscript k is replaced by the subscript j that takes on all values from $j = 1$ to $j = n_i$, and b) Y is replaced by the class mean $\bar{Y}_i$ which is equal to

$$\bar{Y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}. \quad \dots(3)$$

This gives

$$\sum_{j=1}^{n_i} (Y_{ij} - A)^2 = \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 + n_i(\bar{Y}_i - A)^2.$$

If in this relation A is made equal to the general mean $\bar{Y}$, then

$$\sum_{j=1}^{n_i} (Y_{ij} - \bar{Y})^2 = \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 + n_i(\bar{Y}_i - \bar{Y})^2. \quad \dots(4)$$

Finally, the summation with respect to all h classes leads to the fundamental formula of decomposition of the sum of squares:

$$\sum_{i=1}^h \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y})^2 = \sum_{i=1}^h \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 + \sum_{i=1}^h (\bar{Y}_i - \bar{Y})^2 n_i. \quad \dots(5)$$

A particularly important feature of this expression is that a close inspection of its three terms reveals that each of them (provided we take into account the appropriate number of degrees of freedom) can lead to an estimate of the variance of the sample. For the calculation of an estimate, the term must be divided by its own specific number of degrees of freedom. The estimates are all unbiased, i.e., they all converge to the same value if the sample size grows to infinity.

The expressions for the estimates are presented below. The symbols for the estimates of variances are s^2 , s'^2 , s''^2 and the corresponding degrees of freedom are denoted by v , v' , v'' , respectively.

$$s^2 = \frac{1}{N-1} \sum_{ij} (Y_{ij} - \bar{Y})^2 ; \quad v = N-1. \quad \dots(6a)$$

$$s'^2 = \frac{1}{N-h} \sum_{ij} (Y_{ij} - \bar{Y}_i)^2 ; \quad v' = N-h. \quad \dots(6b)$$

$$s'^2 = \frac{1}{h-1} \sum_{ij} n_{ij} (Y_{ij} - \bar{Y})^2 ; \quad v'' = h-1. \quad \dots(6c)$$

The foundation of the analysis of variance can now be outlined as follows. If the universe of elements (i.e., a very large number of observations) from which a sample of 144 elements was drawn was homogeneous then any classification according to any external formal criterion would not introduce any statistical bias and the three estimates of the variance would not be significantly different from each other. If, however, the estimates s'^2 and s''^2 are significantly different, then this difference points to the existence of a systematic reason, i.e., to a real "cause of variation" (term adopted by statisticians). Those causes which disturb the value of s'^2 are termed "causes of variation within classes", for instance variations within the criterion of classification "manufacturer". The causes affecting s''^2 are termed "causes of variation between class means".

In mathematical statistics, the most commonly used ratio of variances is s'^2/s''^2 or its inverse. It is designated by the symbol F, to honour R.A. Fisher who created and fully developed the analysis of variance. The definition of F is therefore the following:

$$\begin{aligned} \text{if } s'^2 > s''^2, \text{ then } F &= \frac{s'^2}{s''^2}, \\ \text{if } s'^2 < s''^2, \text{ then } F &= \frac{s''^2}{s'^2}. \end{aligned} \quad \dots(7)$$

In most of the text books on statistics the F tables are presented in an abridged form. A more complete form may be found in CRC-Standard Math Tables. In the present case, there are $h = 4$ classes, hence $v_1 = 4-1 = 3$; the value of v_2 is $v_2 = N-4 = 144-4 = 140$. In the F table, for $v_1 = 3$, $v_2 = 140$, there are two values of F: an upper value so-called "5% point" which is smaller than the lower value termed "1% point" (often printed in bold-face

type). With respect to these "points", here equal to 2,4 and 3.4 respectively, the numerical values of F can be

- a) below the 5% point,
- b) above the 1% point,
- c) between these two points.

Suppose that the hypothesis that is tested is that the variances s'^2 and s''^2 correspond to samples that are drawn from a completely homogeneous, normally distributed (or near-normally distributed) universe. The upper value a) is that value which in repeated sampling would be exceeded with a probability equal to 0.05 (i.e., 5%). The lower value b) would be, on the other hand, exceeded with a probability of 0.01 (i.e., 1%). If the value of F is below the "5% point" then this value can be considered as not significant and this would be favourable to the formulated hypothesis, i.e., to assuming the homogeneity of the universe.

This section can be summarized colloquially as follows: if the sample is classified according to the criterion "manufacturer" and if such a classification has no systematic effect (or, what is the same, the meters of one manufacturer are not distinguishable from those of other manufacturers) the variances s'^2 and s''^2 are likely to be close to each other. A similar statement can be made concerning the criterion direction. If F is large, i.e., if s'^2 is actually very different from s''^2 then the criterion has a significant effect on the behaviour of the meters. As it has been said above, the convenient points are those which are termed 5% point and 1% point. The values that are located between these points cannot be securely assessed and may require some other investigations.

It must be finally and strongly emphasized, as the conclusion of this theoretical outline of the principles of the analysis of variance, that this method, however powerful (for testing the likelihood of an assumption) does not suggest for what real reason the hypothesis should be accepted or

rejected. It is left to the experience and the imagination of the observer-calculator to uncover the reasons which produce the observed effects.

3.

NUMERICAL CALCULATIONS

The form of equations (6) is not very well adapted to numerical calculations. To illustrate how these calculations should be performed, the results of the observations made in November 1972 will be treated in detail. The principal steps of the calculations are also reproduced, in a more condensed form, in the "outputs" of the computer, the meaning of the symbols used being everywhere the same. The calculations of all other years are then presented only under the form of the computer "outputs". Appendix I.

The 1972 observations (TRE in % form) are displayed in Table IA. They are classified vertically according to the criterion "manufacturer": General Electric (GE), Ferranti-Packard (FP), Westinghouse (W), Sangamo (S). The same table can also be used horizontally, the classification being made according to the criterion "direction": South (S), East (E), North (N), West (W). The calculations will be described for the vertical classification. Their sequence is the following:

$$a) \text{ Class totals } t_i = \sum_{j=1}^{36} Y_{ij} . \quad \text{This gives } t_1, t_2, t_3, t_4. \quad \dots(8)$$

$$\text{Grand total } t_0 : t_0 = t_1 + t_2 + t_3 + t_4.$$

$$\text{Squares: } t_1^2, t_2^2, t_3^2, t_4^2 \text{ and } t_0^2 = (t_1 + t_2 + t_3 + t_4)^2$$

b) General "sum of squares" S:

$$\sum_{k=1}^{144} Y_k^2 = \sum_{i=1}^4 \sum_{j=1}^{36} Y_{ij}^2 ; k = 1, 2, \dots, 144 \quad \left\{ \begin{array}{l} i = 1, 2, 3, 4 \\ j = 1, 2, \dots, 36 \end{array} \right. \quad \dots(9)$$

$$c) \text{ Class means squares } \mu_i : \mu_1 = \frac{t_1^2}{n_1}, \mu_2 = \frac{t_2^2}{36}, \dots, \mu_4 = \frac{t_4^2}{36}. \quad \dots(10)$$

$$d) \text{ Total class mean square } \mu_0 : \mu_0 = \sum_i \mu_i = \mu_1 + \mu_2 + \mu_3 + \mu_4. \quad \dots(11)$$

$$e) \text{ General mean square } M : M = \frac{t_0^2}{N} = \frac{(t_1 + t_2 + t_3 + t_4)^2}{144} \dots(12)$$

The equation (6a) can be, in conformity with (2), put under the form

$$\begin{aligned} \sum_{ij} (Y_{ij} - \bar{Y})^2 &= \sum_{ij} Y_{ij}^2 - N\bar{Y}^2 = S - N\left(\frac{t_0}{N}\right)^2 \\ &= S - \frac{t_0^2}{N} = S - M = [A] \text{ (output)} \end{aligned} \dots(13)$$

This is also valid for (6b) except that the symbol M must be replaced by μ_0 :

$$\sum_{ij} (Y_{ij} - \bar{Y}_i)^2 = S - \mu_0 = [B] \text{ (output).}$$

Finally, the relation (6c) leads to

$$\sum_i (\bar{Y}_i - \bar{Y})^2 = \mu_0 - M = [C] \text{ (output)}$$

The numerical values in the above relations are

$$\begin{aligned} \mu_1 &= 4.71, \mu_2 = 2.89, \mu_3 = 2.49, \mu_4 = 25.23, \mu_0 = 35.32. \\ S &= 58.23. M = 27.41. \\ [A] &= 30.82, [B] = 22.91, [C] = 7.91. \end{aligned}$$

Table of final formulae

$$\begin{aligned} s^2 &= \frac{1}{N-1} \sum_{ij} (Y_{ij} - \bar{Y})^2 = \frac{[A]}{N-1} = \frac{30.82}{143} = 0.22. \\ s'^2 &= \frac{1}{N-h} \sum_{ij} (Y_{ij} - \bar{Y}_i)^2 = \frac{[B]}{N-h} = \frac{22.91}{140} = 0.16. \end{aligned}$$

$$s'^2 = \frac{1}{h-1} \sum_i n_i (\bar{Y}_i - \bar{Y})^2 = \frac{[C]}{h-1} = \frac{7.91}{3} = 2.64.$$

Fisher's Index F:

$$F = \frac{s'^2}{s^2} = \frac{2.64}{0.16} = 16.2$$

According to the meaning of F, this shows that the value of F is very significant as, in the table, the 5% point is equal to 2.4 (for $v_1 = 140$, $v_2 = 3$). This is interpreted as being strongly against the assumption that the sample is homogenous with respect to manufacturers.

The above calculations are also summarized in the upper half of Table IB. The lower half shows the results of the calculations concerning the criterion "direction". Here, $F = 1.56$ which, with respect to 1% point (3.4) is very small. Hence, the assumption that the sample is homogeneous with respect to the geographical orientation of the meters is very likely to be correct.

The examination of all eight tables (I to VIII) confirms the conclusion yielded by the year 1972. The Reader is now justified to ask whether similar conclusions could be reached by our examining directly the numerical data presented in the tables. Consider the values of class totals $t_i (i = 1, 2, 3, 4)$:

	t_1	t_2	t_3	t_4
Manuf.	13.02	10.20	9.46	30.14
Direct.	13.62	14.09	18.35	16.76

The total "Manuf. $t_4 = 30.14$ " is very different from other totals on the same horizontal line but no such remark can be made concerning the second line (Direct.). Thus by classifying meters according to the criterion "manufacturer" we form a group of 36 meters (with $t_4 = 30.14$) which is

substantially different from the other three analogous groups. On the other hand, the classification according to directions produces four groups that are very similar to each other (second line in the table above).

A few conclusions of a general nature which result from the analysis of variance are joined to those which will result from the analysis performed in the next section.

4.

OTHER STATISTICAL INVESTIGATIONS

In the "analysis of variance" a sample is represented by its mean and variance. No attention is paid to its actual internal frequency distribution. However, it is necessary to check that this distribution is not too far from normality (bell-shaped) as the establishment of Fisher's Tables is founded on the postulate that the distribution of the universe is normal or close to normal. The Reader can easily check that such is the case by examining the data presented in the computer "outputs" (Tables I and VIII). In the present section, the internal structure of the samples will play an important role, particularly in relation with the structure of the mixtures of partial subsamples into which the total populations can be subdivided.

Before performing this analysis, let us first emphasize that a sample of 144 elements is actually rather a "small" sample. If it is subdivided into parts, these parts are so small that statistical calculations can lead only to unreliable conclusions. In order to operate on larger samples, all observations made during the eight consecutive years were lumped into one single sample. In addition, it was decided to recalculate all TRE anew but on a six-month basis. This has led to a total population of 2252 elements which, because of eliminations, is a little smaller than the theoretical maximum of $2 \times 144 \times 8 = 2304$. The loss of 52 TRE is due to the elimination of 7 meters (indicated by asterisks in the outputs). The total population consists therefore of:

- a) Eight "May Tables" (analogous to Table IA) containing
1128 TRE calculated for six-month periods, November to May;

- b) Eight "November Tables" containing 1124 TRE calculated for six-month periods, May to November.

The errors recorded in May belong to the period that contains winter months so that this period is called "cold season" and is designated by the subscript "c". The period extending from May to November contains summer months and is called "warm season" (subscript "w"). The individual values of the 2252 TRE are not given in the present memo; however, they figure in classified form in both Fig. 1 and Fig. 2: in Fig. 1 the total (2252) is considered as a "mixture" of two subsamples, subsample "cold season" and subsample "warm season"; in Fig. 2 the total is a mixture of two subsamples, "cold direction" and "warm direction". The statistics of the subdivision into "seasons" are the following:

Cold season (Nov-May): $N_c = 1128$, $\bar{m}_c = 0.57$, st. dev. $s_c = 0.52$

Warm season (May-Nov): $N_w = 1124$, $\bar{m}_w = 0.22$ " " $s_w = 0.59$

Total: $N_t = 2252$, $\bar{m}_t = 0.39$, " " $s_b = 0.58$

Before drawing conclusions from Fig. 1 and 2 it is necessary to mention again, the important fact that, in the course of the eight-year long experiment, seven meters have been eliminated for malfunctioning. The manner in which the registrations have been recorded (the control is not continuous) does not provide a better insight on the manner in which malfunctionings have actually started to appear. In about half of the cases, the meter was found to have simply stopped (and remained stopped for the rest of the experiment). In other cases, the reading was found to be totally erroneous. Only an experienced manufacturer and an expert in electrical measurements can judge the quality of the meters that were tested in the manner described in this memorandum. If the cases of malfunctioning are considered as not significantly numerous, one can say that the meters have been working satisfactorily. A meter seems to be, within the accepted tolerance limits, a reliable instrument in spite of the hard climate of the Ottawa region.

Concerning frequency distribution, it is possible to make the following general remark: all five observed diagrams are reasonably regular, symmetric and bell-shaped but four of them are not really well represented by Gaussian curves. They are leptokurtic with different degrees of leptokurtosis.*³ In the modern statistical theory of errors the presence of leptokurtosis is simply treated as an additional characteristic of a sample without any specific meaning so far as the quality of observations is concerned.

The general impression (as it may result from a superficial look) produced by Fig. 1 is that of "dispersion" while that produced by Fig. 2 is that of "compactness". This suggests the idea that the criterion "season" influences the behaviour of meters more strongly than the criterion "direction". A closer examination shows that in both figures the means with subscript w are close to 0 than the means with subscript c: it seems therefore that in warmer conditions (South and Summer) the registration errors have the tendency to decrease while in colder conditions there is a tendency in the opposite sense. In addition, the means of the subsamples are farther apart in Fig. 1 than in Fig. 2 and the difference in the former is not negligible:

$$\bar{m}_c - \bar{m}_w = 0.35$$

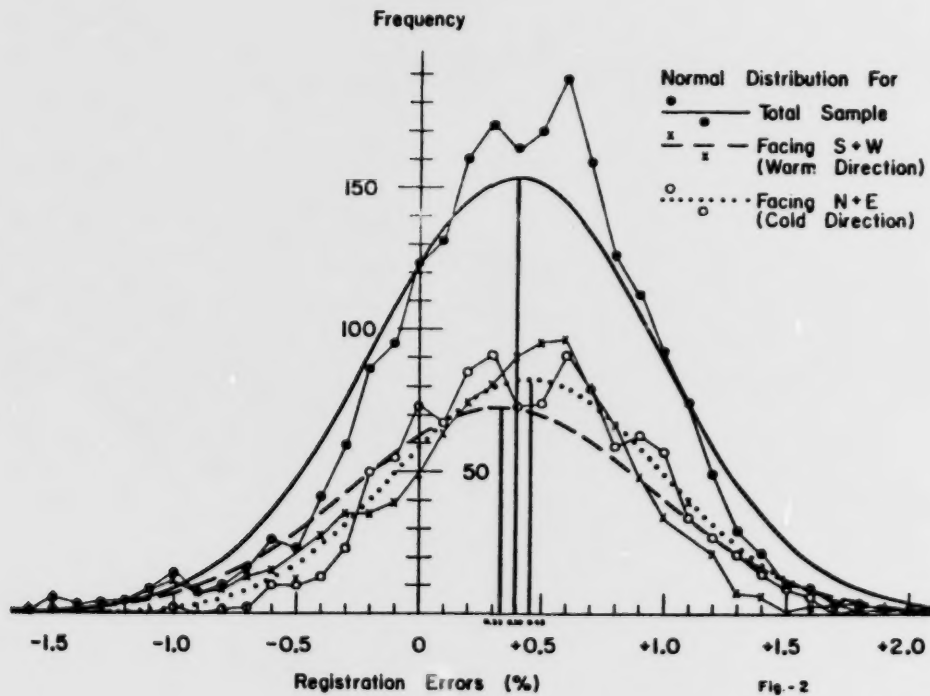
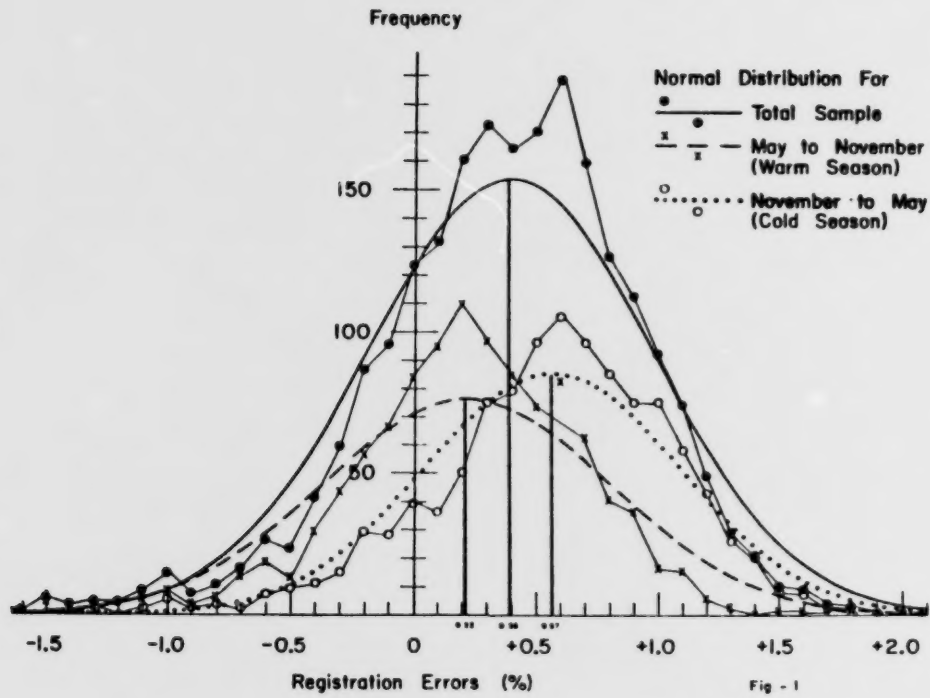
As s_c and s_w are both of the order of 0.55, the difference is of the order of 0.6 standard deviation. This is not an insignificant difference. This can be expressed colloquially as follows: it is to the advantage of the consumer to place his meter on the Southern side of the building and to use

*³ An observed diagram of class frequencies is said to be leptokurtic (K. Pearson's term) if, in its central portion, the observed class frequencies (represented by dots . in both "totals") are greater than the corresponding theoretical class frequencies of the Gaussian curve that fits into the diagram. In the curve representing the total sample (2252 elements) there are nine dots . that are above the Gaussian curve. This central excess is compensated by the corresponding lowering of . dots on both sides of the central part.

the electricity mostly during the warm part of the year. However, his gain would be only "probable" and would be only very slight!

Note

The sample of Total Registration Errors (TRE) analyzed above is also used to illustrate the theory of "mixtures", particularly when the components are of the modulated normal type. The article on this subject is published in the Bulletin Géodésique, 53, No. 1, 1983, p. 62 and can be obtained from the authors (MR or EG) at NRC, Division of Physics, Ottawa, Ontario, Canada K1A 0R6.





APPENDIX I

This appendix contains the tables of the observed data and the computer "outputs" for each of the years 1972 - 1979. In these tables the observed data are represented as in Table IA.

The reader will notice that the numerical results, although somewhat different from year to year, lead to the same conclusions as those presented at the end of Section 2.

Date of Data is 1972 NOV

Date of Data is 1972 NOV

TABLE I-B

	GE	FP	W	S
SOUTH	.01 .20 .12 .39 .07 .17 .16 .01 .54	.72 .56 .95 .83 .24 .40 .44 .91 .49	.62 .60 .09 .76 .14 .72 .73 1.06 .87	1.24 1.48 .72 1.15 .80 .62 .64 1.02 .95
EAST	.31 .30 .11 .65 .36 .58 .56 .65 .47	.13 .27 .43 .34 .07 .07 .40 .53 .05	.05 .29 .21 .38 1.17 .60 .14 .18 .24	.70 .64 1.20 .57 .89 .75 .92 .87 .81
NORTH	.32 .11 .07 .65 .72 .69 .68 .62 .35	.26 .69 .98 .55 .88 .30 .62 .31 .42	.39 .07 .68 .18 .59 .83 .98 .31 .03	.51 .60 .98 .44 .72 .25 .92 .81 1.20
WEST	.22 .46 .36 .05 .42 .99 .40 .40 .63	.50 .66 .30 .77 .03 .50 .04 .20 .32	.17 .40 .03 .28 .15 .43 .56 .57 .08	.27 2.50 .59 .45 .47 .56 1.65 .88 .45
	t_1	t_2	t_3	t_4

Criterion- MANUFACTURER

N= 144.00
h=4

Manufacturer
Totals (t)
Totals squared
Mu

FP
10.20
104.84

W
9.46
89.49

S
30.14
908.42

N-h= 140.00
h-1=3

t zero = 62.82
t zero squar = 3946.35
Mu zero = 35.32
A = 30.82
B = 22.91
C = 7.91
S' square = .16
S' square = 2.64
F = 16.12
(v' 3, v' = 140.00)

Criterion- DIRECTION

N= 144.00
h=4

S= 58.23
M= 27.41

Direction
Totals (t)
Totals squared
Mu

FAST
14.09
198.53
5.51

NORTH
18.35
336.72
9.55

WEST
16.76
280.90
7.80

N-h= 140.00
h-1=3

t zero = 62.82
t zero squar = 3946.35
Mu zero = 37.82
A = 30.82
B = 30.40
C = .42
S' square = .22
S' square = .14
F = 1.56
(v' 3, v' = 140.00)

TABLE III

Date of Data is 1974 NOV

Date of Data is 1974 NOV		Criterion - MANUFACTURER	
S = 47.27 M = 16.67		N = 143.00 h = 4	
Manufacturer		FP	
Totals (t)		6.99	
Totals squared		48.86	
Mu		1.36	
GE		W	
7.31		17.22	
53.44		149.33	
1.53		4.15	
S		25.30	
497.29		13.81	

N = 139.00 h = 3	
t zero = 48.82	
t zero squar = 2383.39	
Mu zero = 20.85	
A = 30.60	
B = 26.42	
C = 4.18	
S' square = .19	
S' square = 1.39	
F = 7.33	
(v' 3, v' 1 = 139.00)	

Criterion - DIRECTION

N = 143.00
h = 4

S = 47.27 M = 16.67	
Direction	
Totals (t)	
Totals squared	
Mu	
SOUTH	
7.97	
63.52	
1.76	

EAST	
13.13	
172.40	
4.79	

NORTH	
16.30	
265.69	
7.38	

WEST	
11.42	
130.42	
3.73	

N = 139.00
h = 3

t zero = 48.82	
t zero squar = 2383.39	
Mu zero = 17.66	
A = 30.60	
B = 29.61	
C = .99	
S' square = .21	
S' square = .33	
F = .64	
(v' 3, v' 1 = 139.00)	

Date of Data is 1974 NOV

.10	.60	.88	.84
.27	.29	.63	.56
.03	-1.26	0.00	.56
.35	.42	.64	.81
.13	.40	.06	.80
.48	.81	1.02	.47
.05	.34	.51	.56
.61	-1.20	.08	.85
.44	-.76	.63	.77

.04	.31	.15	.99
.19	.29	.01	.55
.06	.50	-.07	1.09
.57	.23	.28	.43
.30	-.02	1.38	.79
.17	.83	.54	.63
.52	.38	-.13	.90
.57	.46	-.26	.87
.60	-1.01	.25	.70

.19	.09	.28	.39
0.00	.81	-.18	.38
-.06	.86	.46	.86
.53	.50	-.12	.39
.61	.76	.61	.60
.60	.19	-.16	.13
.50	.61	1.00	.84
.42	.15	1.11	.95
.23	.43	.12	1.22

.06	.29	.26	.22
.32	1.22	.44	.50
.22	1.03	-.18	.43
-.10	.65	.16	.29
***	-.23	0.00	.30
.87	.26	-.46	.39
.25	-.25	1.46	.15
.21	-.08	.49	.72
.50	.33	.33	.37

TABLE IV

Date of Data is 1975 NOV		Date of Data is 1975 NOV		Criterion- MANUFACTURER	
-0.03 -0.19 0.06 0.47 -0.07 -0.42 0.08 -0.05 0.49	0.76 0.21 -0.16 0.37 0.58 -0.22 0.29 -0.27 -0.94	0.96 0.72 0.05 0.81 *** 0.34 0.63 1.02 0.62	0.70 0.62 0.59 0.81 1.02 0.47 0.62 0.84 0.78	S= 40.12 h= 16.10 Manufacturer Totals (t) Totals Squared Mu	GE 6.94 48.16 1.38
-0.51 0.24 0.11 0.55 0.31 0.37 0.49 0.53 0.44	0.29 0.42 0.50 0.25 -0.01 0.83 0.41 0.49 -0.14	0.35 0.11 0.09 -1.14 1.55 0.65 -0.02 -1.11 0.19	1.16 0.61 1.09 0.46 0.80 0.67 0.96 0.89 0.83	t zero = 47.64 t zero squar. = 2269.57 Mu zero = 20.93 A = 24.03 B = 19.19 C = 4.83 S' square = .14 S' square = 11.50 F (9, 3, 977) = 137.00	N-h= 137.00 h-1=3
0.13 -0.06 -0.08 0.40 0.49 0.47 0.39 0.36 0.21	0.05 0.74 0.78 0.50 0.79 0.13 0.59 0.14 0.51	0.30 -1.10 0.44 -1.01 0.67 0.15 -1.00 0.26 0.24	0.39 0.31 0.83 0.36 0.61 0.09 0.81 0.90 1.16	S= 40.12 h= 16.10 Direction Totals (t) Totals Squared Mu	SOUTH 11.99 143.76 4.11
0.02 0.28 0.28 0.18 -0.18 *** 0.72 0.22 0.10 0.47	0.28 0.37 0.83 0.46 -0.29 0.05 -0.36 -0.15 -0.02	0.41 0.53 -1.11 0.26 0.03 -0.26 0.44 0.59 ***	0.17 0.53 0.37 0.25 0.3 0.33 0.60 0.66 0.51	t zero = 47.64 t zero squar. = 2269.57 Mu zero = 17.06 A = 24.03 B = 23.06 C = .96 S' square = .17 S' square = 3 F (9, 3, 977) = 137.00	N-h= 137.00 h-1=3
				Criterion- DIRECTION	
				N= 141.00 h=4	
				{AST 15.71 246.80 6.86	
				NORTH 12.95 167.70 4.66	
				WEST 8.99 48.86 1.44	
				S 23.16 536.39 14.90	
				W 10.64 113.21 3.33	
				FP 6.90 47.61 1.32	

TABLE V

Date of Data is 1976 NOV		Date of Data is 1976 NOV		Criterion- MANUFACTURER	
.19 -.05 .18 .56 -.01 -.21 .32 .02 .64	1.05 .37 -.59 .61 .80 -.09 .35 -1.16 -.95	1.23 .74 .23 .86 -.08 .33 .61 1.05 .61	.85 .69 .75 .87 1.07 .61 .79 .95 .92	N= 140.00 h=4	
				S= 58.24 M= 32.45	
				Manufacturer	
				Totals (t)	
				Totals squared	
				Mu	
				GE	
				12.05	
				145.20	
				4.48	
FP					
12.05					
145.20					
4.48					
W					
14.88					
221.43					
6.33					
S					
14.88					
221.43					
6.33					

TABLE VIII

Date of Data is 1979 NOV		Date of Data is 1979 NOV		Criterion - MANUFACTURE	
.09 .14 .21 .27 .34 .41 .48 .55 .62 .69	1.16 .21 .27 .34 .41 .48 .55 .62 .69 .76	1.25 .80 .27 .92 0.00 .27 .75 1.09 .60	.71 .56 .70 .84 .77 .79 .8 .85	S = 63.47 M = 31.70 N = 138.00 h = 4	
				Manufacturer Totals (t) Totals Squared Mu	
				GE 9.86 97.21 3.14	
				FP 10.06 101.28 2.81	
				W 10.13 38.70 2.59	
				S 14.08 84.05 21.92	
.34 .10 .03 .63 .64 .56 .51 .28	.22 .94 .95 .49 .25 .73 .04 .58	.54 .15 .73 .08 .81 .17 1.09 .50 .32	.58 .47 1.11 .43 .69 .25 .98 1.21 1.32	S = 63.47 M = 31.70 N = 138.00 h = 4	
				Direction Totals (t) Totals Squared Mu	
				SOUTH 12.69 161.04 4.60	
				EAST 18.95 359.10 10.56	
				NORTH 19.80 392.04 11.20	
				WEST 14.70 211.04 5.56	
.12 .36 .25 .06 .88 .93 .36 .28 .69	.54 .71 1.38 .66 .28 .29 .48 .24 .03	.61 .90 .19 .50 .38 .03 .43 .70 .88	.34 .51 .68 .41 .58 .51 .89 .85 .85	S = 63.47 M = 31.70 N = 138.00 h = 4	
				Direction Totals (t) Totals Squared Mu	
				SOUTH 12.69 161.04 4.60	
				EAST 18.95 359.10 10.56	
				NORTH 19.80 392.04 11.20	
				WEST 14.70 211.04 5.56	



APPENDIX II

Note on the correlation between the Total Registration Errors (TRE) and the Disk Errors (DE) in the Project "Meters on the Roof".

This note has been prepared in coauthorship with B. Fulsom, Head, Electrical Laboratory and J. Samuel, summer student (1982).

**DISC ERRORS IN THE PROJECT
"METERS ON THE ROOF"**

I. Introduction

The basic feature of "Total Registration Errors" (TRE) is that they require that the measurements be made at long intervals, at least several months. This implies the consumption of large quantities of electrical energy and the maintenance, during long periods, of appropriate circuitry and installations. One of the reasons for such a situation is that the mechanism of a meter gives reliable readings only when relatively large quantities of watt hours are considered. Nevertheless, for the manufacturer and for the user who consumes large amounts of electrical energy, TRE are obviously of the greatest importance and must be investigated.

In current practice, however, one of the routine operations is the determination of the meter's error over a very short time e.g. only a few minutes. It is obvious that such a determination cannot be as reliable as that of the total registration error but, for obvious reasons, it must be performed. When a user finds that a meter requires to be checked, it is transferred to an appropriate laboratory in which the so-called Disk Error (DE) is determined.

The fundamental property of an induction type meter is that a turn of its rotating disk corresponds to a constant number of units of energy (joule or watt hour). This constant, denoted by the symbol "Kh", is somewhat disturbed by the conditions in which the measurement is performed. The determination of the TRE benefits from the averaging effect due to the very large number of turns which take place during long periods of time.

Another important point that must be mentioned concerns the current used for DE determinations. The influence of its intensity on Kh is not simple and

in any case, is far from linear. In common practice, K_h is determined with the current being equal to 2.5 A and 25 A. All manufacturers aim at producing meters that possess prescribed values for disk constants and a meter is considered (by the manufacturer) as acceptable for domestic electrical installation if its Disk Error is not larger than 1%. In fact, such a limit is narrower than the legal tolerance (2%) because the checking of new meters is based on sampling procedures and not on the examination of every individual instrument.

The reference for all DE is an instrument called "Transfer standard". It generates pulses, the time between two consecutive pulses corresponding to a constant amount of electrical energy. It is used in the laboratory of Legal Metrology to determine the constant K_h of the "Reference Watt hour Standard". The constant K_h is adjusted so that one watt hour corresponds to three revolutions of the disk. In other words, $K_h = 1/3$ and legally a meter is judged acceptable if it is not in error by more than 2%.

II. Experimental Work and Calculation of Correlation:

It is obvious that even on a high quality meter a linear relationship between its TRE and DE will not be absolutely perfect. In statistical language we can however expect that these two kinds of errors will be, although not "perfectly", but "very highly correlated". In the course of the eight-year long experiment already described, a certain number of determinations of the Disk Errors have been also performed. However, for practical reasons, the population of DE thus constituted could not be treated in its entirety and only three separate "periods" could be calculated and statistically analysed. The first and the third periods extend over a year, the second period extends over two consecutive years. The DE for a period is the average of the observations performed at the beginning, the middle and the end of the period. This average is denoted by the symbol X_i so that, in each period, to a TRE Y_i corresponds a Disk Error X_i . The tables of the X_i are not presented

in this memorandum and the data concerning the periods are summarized as follows:

First period: 1 Nov 1971 to 11 Oct 1972, N = 144 meters.
Second " : 15 Oct 1974 to 15 Oct 1976, N = 139 meters.
Third " : 12 Oct 1975 to 15 Oct 1979, N = 138 meters.

The sequence of operations leading to the value of the "correlation index" r can be presented as follows:

A) Calculation of means $\bar{Y}$ and $\bar{X}$:

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^N \bar{Y}_i ; \bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$$

B) Calculation of the deviations y_i and x_i :

$$y_i = Y_i - \bar{Y} ; x_i = X_i - \bar{X}.$$

C) Calculation of the variances s_y^2 and s_x^2 :

$$s_y^2 = \frac{1}{N} \sum_{i=1}^N y_i^2 ; s_x^2 = \frac{1}{N} \sum_{i=1}^N x_i^2.$$

D) Calculation of the covariance:

$$\mu_{11} = \frac{1}{N} \sum_{i=1}^N y_i x_i.$$

E) The correlation index r, as it results from these values, is:

$$r = \frac{\mu_{11}}{s_y s_x}$$

F) The significance of the obtained numerical value of r is obtained by referring this value to the corresponding Student's t -variate. The relation between t and r is:

$$t = \frac{r}{1-r^2} (N - 2)$$

A strong correlation between y and x (as it is expected here) leads to values of r that are close to unity; that in its turn leads to large values of t . With, for instance, $r = 0.5$ ($N = 144$), $t = 6.9$. This is a very large t and it indicates the existence of a strong correlation.

The numerical values of various quantities mentioned above are, in present analysis, given in the following table.

FIRST PERIOD (1972)	SECOND PERIOD (1976)	THIRD PERIOD
$\mu_{11} = 0.170$	0.180	0.250
$s_y^2 = 0.215$	0.185	0.236
$s_x = 0.464$	0.430	0.486
$s_x^2 = 0.301$	0.324	0.361
$s_x = 0.549$	0.569	0.601
$r = 0.667$	0.736	0.855

It is to be noted that all values of r are larger than 0.5. For $r = 0.667$, t is larger than 10. The correlations between X and Y in all three periods are therefore extremely strong. The overall conclusion of the experiment and of its statistical analysis is that there is no reason to invalidate the traditional use of the Disk Error test to check the correctness of an electrical meter.

If the circumstances permit (i.e. if the appropriate manpower becomes available) it may become possible to apply the Analysis of Variance directly to Disk Errors. It would be interesting to see how this analysis agrees, in its conclusions, with that of the Analysis of Variance applied to the Total Registration Errors (TRE).

DOCUMENT DE TRAVAIL

Compteurs sur le toit

Examen statistique
des résultats obtenus

CONSOMMATION ET
CORPORATIONS CANADA
MÉTROLOGIE LÉGALE

A-MON-1984/58; F; C
(Métrologie Légale)

Canada



Government
of Canada

Gouvernement
du Canada

MEMORANDUM

NOTE DE SERVICE

TO
A

Corinne MacLaurin
Bibliothécaire en chef/
Chief Librarian

FROM
DE

Directeur
Direction de la Métrologie légale

Director
Legal Metrology Branch

SUBJECT
OBJET

COMPTEURS SUR LE TOIT

Veillez trouver ci-joint une copie des versions française et anglaise du document intitulé Compteurs sur le toit préparé par MM. M. Romanowski et E. Green pour le compte de la Direction de la Métrologie légale de Consommation et Corporations Canada.

Ce "document de travail" présente une analyse statistique des résultats d'une expérience qui s'est échelonnée sur plus d'une dizaine d'années et qui a pris fin il y a bientôt cinq ans. Ses conclusions demeurent cependant très actuelles et sont susceptibles d'intéresser particulièrement les statisticiens.

Je me permets de conseiller aux lecteurs qui s'intéressent aux analyses de variances de lire la Note à la page 24 du document.

Dans le cas où vous désireriez d'autres copies de ce document, vous êtes priés d'en faire la demande auprès de Hélène Pagé au numéro de téléphone (613) 990-8617.

SECURITY - CLASSIFICATION - DE SÉCURITÉ

OUR FILE/NOTRE RÉFÉRENCE

G6520-5

YOUR FILE/VOTRE RÉFÉRENCE

DATE

84-10-19

METERS ON THE ROOF

Please find enclosed copies of the English and French versions of the publication entitled Meters on the Roof prepared by M. Romanowski and E. Green for the Legal Metrology Branch of Consumer and Corporate Affairs Canada.

This "working paper" contains statistical analysis of variance on data from an experiment with electricity meters that covered more than a decade and was stopped almost five years ago. Nevertheless, its conclusions will be of interest, particularly to statisticians.

Those readers who are interested in analysis of variance should read the Note on page 22.

If additional copies of this document are required, please contact Hélène Pagé at (613) 990-8617.

Richard G. Knapp

p.j.

encl.

CCA

TK 301
R6F

Mémoire technique

Examen statistique des résultats obtenus

dans le projet "Compteurs sur le toit".

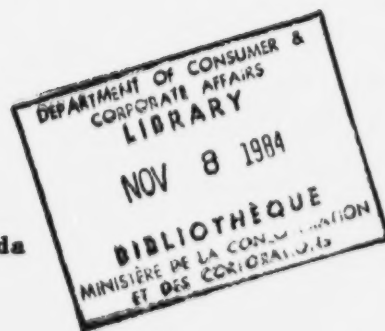
M. Romanowski

et

E. Green

Métrologie légale

Consommation et Corporations Canada



**L'analyse et les conclusions contenues dans ce rapport ne représentent pas
nécessairement le point de vue du ministère.**

© Ministre des Approvisionnement et Services Canada 1984

N° de cat. RG 23-80/1984F

ISBN 0-662-92723-0

Available in English

REMERCIEMENTS

Les auteurs expriment leur gratitude au directeur, M. R.G. Knapp, pour son aide et ses encouragements constants tout au long de la préparation du présent mémoire.

Ils tiennent aussi à remercier LEXICA INC. qui a assuré avec compétence la dactylographie des versions anglaise et française du manuscrit.

TABLE DES MATIÈRES

Préface	1
Description du projet	5
Description de l'analyse de la variance	11
Calculs numériques	17
Autres recherches statistiques	21
Appendice I	
Tableau I-B à VIII	27
Appendice II	
Les erreurs du disque	37

PRÉFACE

Pour comprendre les raisons qui ont conduit au projet expérimental dont les résultats sont analysés dans le présent rapport, il est utile de souligner les événements qui ont eu lieu au début des années 60 et de revoir les lignes de conduite de la Direction de la métrologie légale en matière de vérification, de scellage et de revérification des compteurs.

La Loi sur l'inspection de l'électricité précisait que tous les appareils de mesure utilisés à des fins commerciales devaient, avant d'être mis en service, être vérifiés en vue d'établir leur exactitude en fonction des limites d'erreur autorisées, puis être scellés par un inspecteur du gouvernement pour les mettre à l'abri de toute falsification. Ces dispositions s'appliquaient aussi, cela va de soi, aux compteurs d'électricité domestiques.

Étant donné qu'on ne peut s'attendre qu'un compteur demeure exact indéfiniment, la Loi prévoyait également que chaque compteur devait être vérifié de nouveau à intervalles fixes. À l'origine, la durée du scellage attestant la vérification avait été fixée, par voie de réglementation, à six ans pour tous les compteurs d'électricité, mais, en ce qui a trait aux compteurs domestiques, cette période a, par la suite, été étendue à huit ans. Ainsi, jusqu'au début des années 60 les services publics avaient l'habitude de retirer du service tous les compteurs dont la période de validité du scellé était expirée, d'effectuer un travail de révision et d'entretien dans un atelier, puis de les présenter à un inspecteur du gouvernement pour qu'il les vérifie et les scelle de nouveau.

Les progrès techniques réalisés dans la conception et la fabrication des compteurs, notamment la mise au point des compteurs à suspension magnétique, ont suscité, tant de la part des services publics que des fabricants, des demandes de prolongation de la durée du scellage, fondées sur l'argument qu'un compteur de qualité supérieure pouvait satisfaire pendant une période plus longue aux tolérances permises.

À la même époque, la Direction de métrologie légale était en train d'élaborer et de parfaire ses plans d'échantillonnage statistique pour la vérification et la revérification des compteurs. Au sein de son effectif, plusieurs ont alors décidé d'entreprendre un certain nombre d'expériences à long terme en vue d'éprouver la validité des solutions en voie d'être adoptées en matière de vérification et d'aider à l'élaboration des lignes directrices de la Direction en ce qui concerne les aspects métrologiques des compteurs watt-heuremètres domestiques. De plus, l'idée qu'un compteur correctement fabriqué pouvait être installé indifféremment à l'intérieur ou à l'extérieur sans que cela n'altère de façon sensible son fonctionnement à long terme, a encouragé l'utilisation généralisée des compteurs à suspension magnétique. Le personnel de la Direction estimait que cette idée, si logique et si attrayante qu'elle fût, devait être vérifiée expérimentalement dans un milieu naturel pendant plusieurs années consécutives. C'est ainsi qu'est né, en 1968, le projet des "Compteurs sur le toit".

Cent quarante-quatre compteurs, soit trente-six provenant de chacun des quatre fabricants de l'époque, furent installés sur une structure construite à cette fin sur le toit de l'immeuble de la Métrologie légale. Avant d'être installé pour le programme d'essai, chaque compteur a fait l'objet d'un étalonnage et d'une mise au point, sur courants de 2,5 et de 25 ampères, pour ramener sa limite d'erreur le plus près possible de zéro. En 1970, l'analyse des observations trimestrielles de la première année complète révélait que les résultats n'étaient pas concluants parce que chacun des trois sous-groupes de compteurs était alimenté par une des phases distinctes d'un système de triphasé. Il a donc été décidé de vérifier les compteurs après les avoir connectés en série sur courant monophasé. On a placé, sur ce même courant, trois compteurs "moniteurs" étalonnés situés dans l'immeuble. Ces compteurs devaient enregistrer, aux fins des calculs statistiques, le nombre considéré "exact" de watts-heures consommés par les compteurs installés sur le toit.

À la suite de cette modification, les observations faites au cours de chaque trimestre de 1971 ont été analysées et décrites dans un rapport technique publié en 1975. Le relevé trimestriel des erreurs d'enregistrement s'est poursuivi de façon ininterrompue jusqu'en octobre 1979, mois qui marquait la fin de la phase expérimentale du projet. Les relevés directs des compteurs,

accumulés au cours des mois de mai et octobre pendant huit années (2 252 observations), ont ensuite été convertis en pourcentage afin d'en faciliter l'analyse.

En 1982, la Direction a décidé de faire une analyse statistique des variances en fonction des critères "fabricant" et "orientation".

En étudiant les résultats de cette analyse des variances, le lecteur sera frappé par la quantité et la diversité des renseignements que les auteurs ont réussi à tirer des données expérimentales. Il a été possible de diviser, d'après des critères précis, les 2 252 échantillons en sous-échantillons bien déterminés. On a pu, par exemple, examiner les observations faites sur tous les compteurs d'un certain fabricant, ou faites sur tous les compteurs orientés vers le nord, ou faites au cours de la saison chaude. L'échantillon se prêtait donc tout à la fois à une analyse de la variance et à une étude de la "théorie des mélanges", cette dernière constituant l'un des principaux sujets de recherche des auteurs du présent rapport (voir la bibliographie).

Pour conclure, il y a lieu de considérer comme une heureuse coïncidence le fait que, pendant que l'analyse de la variance était adaptée au projet "Compteurs sur le toit", la Division de physique du Conseil national de recherches étudiait et élaborait la théorie des mélanges de façon totalement indépendante. La jonction finale de ces deux domaines de la statistique, et le succès qui en est résulté, montre que deux organismes peuvent tirer parti d'une collaboration et d'une aide mutuelle.

La Direction de la métrologie légale a une dette de reconnaissance envers M. E. Green, de la Division de la physique du CNR, et plus particulièrement, envers M. Romanowski, conseiller auprès de la Direction.

Richard G. Knapp

Directeur

Direction de la métrologie légale



1.

DESCRIPTION DU PROJET

Un bref historique du projet, qui contient la description du montage original, est donné dans l'Introduction rédigée par Richard G. Knapp, directeur de la Métrologie légale.

Au point de vue purement statistique, chacune des années de l'expérience, qui a duré pendant huit années consécutives, est représentée par quatre "échantillons" de EET (erreur d'enregistrement totale). Chaque échantillon est constitué par les relevés effectués sur la totalité des 144 compteurs observés. Chacun des quatre manufacturiers (General Electric, Ferranti-Packard, Westinghouse, Sangamo) est représenté par un groupe de 36 compteurs considérés comme étant nominalement identiques entre eux. Un "groupe" est subdivisé en quatre "cellules": l'une d'elles fait face au sud (S), une à l'est (E), une au nord (N) et une à l'ouest (O). Une cellule (9 compteurs) est ainsi le plus petit ensemble où deux compteurs peuvent être interchangeables sans influencer les résultats de l'analyse.

Les enregistrements étaient effectués quatre fois par an: en février, mai, août et novembre. Si aucun des compteurs n'était éliminé, ceci aurait conduit à un total de 4608 EET. En réalité, ce total se trouve quelque peu réduit par le fait que le nombre initial de 144 compteurs a progressivement diminué au cours des années et se trouva finalement réduit à 136.

Une opération d'enregistrement, c.à.d. le relevé des indications de tous les compteurs, représente pour une équipe de deux membres du laboratoire d'électricité deux ou trois journées de travail. La réduction des relevés c.à d. leur conversion en "pourcentage" (%) demande beaucoup de temps. Par

suite du manque de main d'oeuvre (techniciens et calculateurs), au début de notre analyse seulement les résultats des mesures de novembre ont pu être réduits. L'analyse de la variance était donc appliquée aux résultats des huit années consécutives (novembre à novembre), et le milésime, p.ex. 1972 indique (en %) les EET de la période allant de novembre 1971 à novembre 1972. (Voir table IA.)

Après que quelques calculs aient été effectués, il est devenu évident que pour des analyses autres que celle de la variance, des échantillons plus grands seraient plus appropriés. On a alors pris la décision de réduire aussi les relevés effectués en mai. Les calculs décrits dans la section 4 sont ainsi effectués sur des échantillons qui, par consentement général, sont considérés comme "grands": leurs dimensions sont de l'ordre de 2000. Les conclusions qui sont tirées des échantillons de cette grandeur sont considérées comme suffisamment probantes pour conduire à des indications pratiques. Avant d'appliquer les méthodes de l'analyse statistique, examinons d'abord la nature de quelques questions générales qui peuvent être posées et auxquelles l'analyse qui suit peut fournir des réponses plus ou moins complètes.

- a) Tenue générale et fonctionnement des instruments quand ils sont utilisés dans des conditions particulières et la probabilité qu'ils vont se comporter d'une manière similaire quand ils sont ensuite placés dans les conditions d'utilisation normales. Il est évident qu'il est très important, pour le manufacturier, pour les institutions métrologiques et, cela va de soi, pour le consommateur de l'énergie électrique de déterminer quelles conditions atmosphériques peuvent influencer le fonctionnement d'un compteur. Est-il possible de prédire comment la position d'un compteur sur une maison ou une usine peut influencer la somme d'argent que l'utilisateur peut avoir à payer? Le fait qu'un compteur est placé sur le côté ensoleillé d'une maison favorise-t-il le producteur ou le consommateur de l'énergie électrique?



- b) Influence de critère "manufacturier".* Quoique tous les compteurs sont supposés être construits "identiquement", il est évident que chaque manufacturier est susceptible d'introduire quelques détails de construction dont la présence peut influencer les résultats.
- c) Influence du critère "direction géographique". Nous avons ici l'une des raisons majeures du fait que l'expérience a été entreprise. Il est, en effet, impossible d'affirmer à priori qu'un compteur tourné vers le sud va enregistrer l'électricité de la même manière qu'un compteur tourné vers le nord. L'orientation peut agir non seulement par la température mais aussi par les facteurs tels que vents, humidité, etc.
- d) Influence du critère "saison". Ceci pose des problèmes qui sont, du point de vue des calculs statistiques, identiques à ceux posés par le critère "direction".

L'information fournie par l'analyse de ces problèmes est aussi importante par les conséquences sur le plan scientifique et technique que sur le plan économique.

La méthode mathématique utilisée pour comparer le comportement des compteurs en tenant compte d'un certain critère de classification est celle de l'analyse de la variance. Les principes de ce puissant outil mathématique ont été établis par R.A. Fisher et sont, en général, exposés dans tous les manuels de statistique.

* En statistique les termes tels que "manufacturier" "direction" etc. désignent des "critères d'après lesquels un échantillon peut-être classifié.

Pour aider le lecteur, les fondements de l'analyse de la variance, tels qu'ils sont utilisés dans le présent rapport, sont décrits brièvement dans la section suivante.



2.

DESCRIPTION DE L'ANALYSE DE LA VARIANCE

Dans le présent rapport (comme dans tous les travaux précédents sur les sujets analogues) le symbole Y (lettre capitale munie d'indices appropriés) désigne l'erreur d'enregistrement totale (EET) calculée sur une période annuelle c.à.d. de novembre à novembre. Un nombre tel que Y_k (compris entre 1 et 144) indique le "pourcentage" par lequel l'enregistrement du k^{eme} compteur s'écarte de celui des moniteurs*²). Les moniteurs sont des étalons de référence placés dans un laboratoire (non à ciel ouvert). Ils forment la base pour l'évaluation des enregistrements des compteurs étudiés. Si, par ex., $Y_{10} = 0.23$ cela signifie que le compteur de rang 10 a enregistré (après avoir été sous tension pendant douze mois) une quantité d'énergie électrique de 0.23% inférieure à celle enregistré par les moniteurs. Cette dernière est parfois appelée, quoique pas toujours correctement, le "vrai" enregistrement, en kWh.

Pour rendre le présent rapport plus facilement accessible aux non-statisticiens, seulement l'analyse de la variance avec un seul critère sera utilisée: les 144 compteurs seront donc subdivisés en quatre classes ou bien d'après le critère "direction" ou le critère "manufacturier". Au début de l'expérience (1972) toutes les classes étaient de la même grandeur (36 compteurs). Ainsi qu'il a déjà été dit plus haut les grandeurs initiales n'ont pas pu être maintenues car au cours des années, certains compteurs ont été (pour diverses raisons) éliminés mais non remplacés.

* 2 Les moniteurs sont des étalon de référence placés dans un laboratoire et non à l'extérieur. Ils constituent la base pour l'évaluation de l'enregistrement des étalons examinés.

Si un échantillon de N éléments Y_k ($k = 1, 2, \dots, N$) est considéré dans sa forme non classifiée, la moyenne est donnée par la formule

$$\bar{Y} = \frac{1}{N} \sum_{k=1}^N Y_k. \quad \dots(1)$$

La relation la plus importante dans tous les calculs qui suivent est celle qui exprime la somme des carrés des déviations des éléments par rapport à n'importe quel nombre A, en fonction des déviations par rapport à la moyenne $\bar{Y}$ de l'échantillon.

$$\sum_k (Y_k - A)^2 = \sum_k (Y_k - \bar{Y})^2 + N(\bar{Y} - A)^2. \quad \dots(2a)$$

Lorsque $A = 0$, ceci se réduit à

$$\sum_k Y_k^2 = \sum_k (Y_k - \bar{Y})^2 + N\bar{Y}^2. \quad \dots(2b)$$

Si l'échantillon est distribué en "classes" ces classes seront désignées par l'indice i ; le total des éléments de la $i^{\text{ème}}$ classe sera désigné par le symbole n_i . Le rang dans la classe sera indiqué par l'index j de sorte que $j = 1, 2, \dots, n_i$. Un élément aura ainsi deux indices p.ex. Y_{ij} :

$$Y_{ij} \text{ avec } i = 1, 2, \dots, h \text{ (c.à d. } h \text{ classes; ici } h = 4) \\ j = 1, 2, \dots, n_i \text{ (} n_i \text{ éléments dans } i^{\text{ème}} \text{ classe)}$$

La $i^{\text{ème}}$ classe constitue en elle même un échantillon non classifié de sorte qu'il est possible de lui appliquer la relation (1) dans laquelle a) l'indice k est remplacé par l'indice j lequel prend alors toutes les valeurs de $j = 1$ à $j = n_i$ et b) $\bar{Y}$ est remplacé par la moyenne $\bar{Y}_i$ de la classe, laquelle est donc égale à:

$$\bar{Y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}. \quad \dots(3)$$

Ceci donne

$$\sum_{j=1}^{n_1} (Y_{1j} - A)^2 = \sum_{j=1}^{n_1} (Y_{1j} - \bar{Y}_1)^2 + n_1 (\bar{Y}_1 - A)^2.$$

Si dans cette expression on fait A égal à la moyenne générale $\bar{Y}$ on obtient:

$$\sum_{j=1}^{n_1} (Y_{1j} - \bar{Y})^2 = \sum_{j=1}^{n_1} (Y_{1j} - \bar{Y}_1)^2 + n_1 (\bar{Y}_1 - \bar{Y})^2. \quad \dots(4)$$

Finalement, la sommation par rapport à toutes les classes h conduit à la formule fondamentale de la décomposition de la somme des carrés:

$$\sum_{i=1}^h \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y})^2 = \sum_{i=1}^h \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 + \sum_{i=1}^h (\bar{Y}_i - \bar{Y})^2 n_i \quad \dots(5)$$

Une très importante propriété de cette égalité réside dans le fait qu'un examen de ses trois termes montre que chacun d'eux peut conduire à une estimation de la variance de l'échantillon (à condition d'évaluer correctement le nombre des degrés de liberté). Pour le calcul de l'estimation, le terme doit être divisé par son propre nombre de degrés de liberté. Les estimations sont non-biaisées c.à.d. qu'elles convergent toutes vers la même limite lorsque la grandeur de l'échantillon tend vers l'infini.

Les expressions de l'estimation sont données plus bas. Les symboles pour les estimations des variances sont s^2 , s'^2 , s''^2 et les degrés de libertés qui leur correspondent sont v , v' , v'' .

$$s^2 = \frac{1}{N-1} \sum_{ij} (Y_{ij} - \bar{Y})^2 ; \quad v = N-1. \quad \dots(6a)$$

$$s'^2 = \frac{1}{N-h} \sum_{ij} (Y_{ij} - \bar{Y}_i)^2 ; \quad v' = N-h. \quad \dots(6b)$$

$$s''^2 = \frac{1}{h-1} \sum_{ij} n_i (\bar{Y}_i - \bar{Y})^2 ; \quad v'' = h-1. \quad \dots(6c)$$

Les bases de l'analyse de la variance peuvent être présentées comme suit. Si l'univers des éléments (p.ex. un très grand nombre d'observations) dont un échantillon de 144 éléments a été tiré, était homogène, alors aucune classification (d'après un critère formel) ne pourrait introduire une structure quelconque statistiques. De sorte que les trois estimations de la variance ne pourraient être sensiblement différentes l'une de l'autre. Si, cependant, les estimations sont sensiblement différentes, alors cette différence indique qu'il doit exister une raison systématique c.à.d. une véritable cause de variation (terme adopté par les statisticiens). Celles des causes qui perturbent la valeur de s'^2 sont nommées causes des variations à l'intérieur des classes et celles qui perturbent s''^2 sont nommées causes des variations entre les moyennes des classes.

Dans la statistique mathématique, le rapport le plus utilisé est s'^2/s''^2 ou son inverse. Il est représenté par le symbol F en honneur de R.A. Fischer qui a créé et développé l'analyse de la variance. La définition de F est:

$$\begin{aligned} \text{si } s'^2 > s''^2, \text{ alors } F &= \frac{s'^2}{s''^2}, \\ &\dots(7) \\ \text{si } s'^2 < s''^2, \text{ alors } F &= \frac{s''^2}{s'^2}. \end{aligned}$$

Dans la plupart des manuels de statistiques, les tables de F sont sous une forme abrégée. Des tables plus complètes se trouvent dans "CRC-Standard Math. Tables". Dans le présent rapport il y a $h = 4$ classes, par conséquent $v_1 = 4-1 = 3$. La valeur de v_2 est égale à $v_2 = N-4 = 140$. Dans les tables de F, pour $v_1 = 3$ et $v_2 = 140$ il y a deux valeurs de F: la valeur supérieure appelée "point 5%" est plus grande que la valeur inférieure appelée "point 1%" (souvent imprimée en caractères gras). Par rapport à ces deux "points", égaux ici 2.4 et 3.4 respectivement, les valeurs numériques de F peuvent être

- a) inférieures au point 5%,
- b) supérieures au point 1%,
- c) entre ces deux points.

Supposons que l'hypothèse à tester est que les variances s'^2 et s''^2 des échantillons sont tirées d'une population complètement homogène et normalement (ou presque normalement) distribuée. La valeur a) ci-dessus est celle qui dans un échantillonnage répété serait dépassée avec une probabilité égale à 0.05 (c.à.d. 5%). La valeur b) serait, par contre, dépassée avec une probabilité de 0.01 (c.à.d. 1%). Si la valeur de F est en dessous du "point 5%" elle peut être considérée comme insignifiante et ceci serait alors en faveur de l'hypothèse formulée c.à.d. que la population est homogène.

Cette section peut être résumée comme suit:

si un échantillon est classifié d'après le critère "manufacturier" et si cette classification n'a pas d'effet systématique (ou, ce qui revient au même, les compteurs d'un manufacturier ne peuvent être distingués de ceux des autres manufacturiers) les variances s'^2 et s''^2 sont vraisemblablement proches l'une de l'autre. On peut en dire autant au sujet du critère "direction". Si F est grand c.à.d. si s'^2 est en réalité très différent de s''^2 alors le critère a un effet considérable sur le comportement des compteurs. Ainsi qu'il a été dit plus haut, les points de repère commodes sont ceux qui sont nommés "point 5%" et "point 1%". Les points situés entre les deux ne peuvent pas être évalués plus exactement et peuvent demander un complément d'investigation.

Finalement, il faut souligner fortement, en tant que la conclusion de cette esquisse théorique de l'analyse de la variance, que cette méthode, quoique très efficace (pour juger la probabilité d'une hypothèse) ne suggère pas pour quelles raisons réelles une hypothèse doit être retenue ou rejetée. Il reste à l'expérience et à l'imagination de l'observateur - calculateur de découvrir les vraies raisons qui produisent les effets observés.



c) Carrés moyens des classes μ_1 : $\mu_1 = \frac{t_1^2}{n_1}$, $\mu_1 = \frac{t_1^2}{36}$, ... $\mu_4 = \frac{t_4^2}{36}$ (10)

d) Total des carrés moyens μ_0 : $\mu_0 = \sum \mu_1 = \mu_1 + \mu_2 + \mu_3 + \mu_4$ (11)

e) Carré moyen général: $M = \frac{t_0^2}{N} = \frac{(t_1 + t_2 + t_3 + t_4)^2}{144}$... (12)

$$\sum_{ij} (Y_{ij} - \bar{Y})^2 = \sum_{ij} Y_{ij}^2 - N\bar{Y}^2 = S - N\left(\frac{t_0}{N}\right)^2$$

$$= S - \frac{t_0^2}{N} = S - M = [A] \text{ (Sortie)} \quad \dots (13)$$

Ceci s'applique aussi à (6b) excepté que le symbole M est remplacé par μ_0 .

$$\sum_{ij} (Y_{ij} - \bar{Y}_i)^2 = S - \mu_0 = [B] \text{ (Sortie)}$$

Enfin, la relation (6c) conduit à

$$\sum_i n_i (\bar{Y} - \bar{Y})^2 = \mu_0 - M = [C] \text{ (Sortie)}$$

Les valeurs numériques dans les relations ci dessus sont:

$$\mu_1 = 4.71, \mu_2 = 2.89, \mu_3 = 2.49, \mu_4 = 25.23, \mu_0 = 35.32.$$

$$S = 58.23. M = 27.41.$$

$$[A] = 30.82, [B] = 22.91, [C] = 7.91.$$

Table des formules finales:

$$s^2 = \frac{1}{N-1} \sum_{ij} (Y_{ij} - \bar{Y})^2 = \frac{[A]}{N-1} = \frac{30.82}{143} = 0.22.$$

$$s'^2 = \frac{1}{N-h} \sum_{ij} (Y_{ij} - \bar{Y}_i)^2 = \frac{[B]}{N-h} = \frac{22.91}{140} = 0.16.$$

$$s^2 = \frac{1}{h-1} \sum_{i=1}^h n_i (\bar{y}_i - \bar{y})^2 = \frac{[C]}{h-1} = \frac{7.91}{3} = 2.64.$$

L'index de Fisher F est:

$$F = \frac{s^2}{s'^2} = \frac{2.64}{0.16} = 16.2$$

D'après le sens de l'index F, ceci montre que la valeur de F est très significative car, d'après les tables, le "point 5%" est égal à 2.4 (pour $v_1 = 140$, $v_2 = 3$). Cette valeur est donc considérée comme étant très défavorable à l'hypothèse que l'échantillon est homogène par rapport au critère "manufacturier".

Les calculs ci-dessus sont aussi résumés dans la partie supérieure du Tableau IB. La partie inférieure montre les résultats des calculs concernant le critère "direction". Ici, $F = 1.56$ ce qui est très petit par rapport au "point 1%" (3.4): l'hypothèse que l'échantillon est homogène par rapport à l'orientation géographique a donc beaucoup de chance d'être correcte.

L'examen des huit tables (I à VIII) confirme la conclusion tirée de l'année 1972. Le lecteur est autorisé à se demander si une conclusion similaire pourrait être tirée d'un examen direct des données numériques contenues dans les tables. Considérons les valeurs des totaux

	t_1	t_2	t_3	t_4
Manuf.	13.02	10.20	9.46	30.14
Direct.	13.62	14.09	18.35	16.76

Le total t_4 (Manuf), égal à 30.14, est très différent de ceux qui sont situés sur la même ligne horizontale; on ne peut pas en dire autant de la seconde ligne horizontale (Directions). Ainsi, en classant les compteurs d'après le critère "manufacturier" on trouve, parmi les quatre classes de 36 compteurs, une classe où $t_4 = 30.14$. Ce total est donc fortement différent des trois autres totaux. Par contre, la classification d'après les "directions" conduit

à quatre classes (de 36 chacun) qui sont très semblables l'une à l'autre (seconde ligne du tableau).

Les conclusions d'ordre général auxquelles conduit l'analyse de la variance sont amalgamées avec celles qui résultent de l'analyse effectuée dans la section suivante.

4.

AUTRES RECHERCHES STATISTIQUES

Dans l'analyse de la variance, un échantillon est représenté par sa moyenne et sa variance. On ne tient pas compte de la distribution intérieure de ses fréquences. Il faut, tout de même, vérifier que cette distribution n'est pas trop loin de la normalité (courbe en cloche): le fondement même des tables de Fisher est basé sur le postulat que cette distribution est normale ou presque normale. Le lecteur peut aisément vérifier que tel est le cas en examinant les Tables I à VIII. Dans la présente section, c'est la structure intérieure des échantillons qui va jouer un rôle important, en particulier en relation avec la structure des "mélanges" des sous-échantillons dans lesquels la population totale peut être subdivisée.

Avant de commencer cette analyse, il faut d'abord souligner le fait qu'un échantillon de 144 éléments est, à vrai dire, un "petit" échantillon. Si on le subdivise encore, les parties deviennent si petites que les calculs statistiques conduisent à des conclusions fort incertaines. Afin de pouvoir manipuler des échantillons plus considérables, les observations faites pendant les huit années consécutives ont été jointes afin de ne former qu'un seul échantillon. En plus il a été décidé de recalculer toutes les EER mais sur la base demi-annuelle. Ceci a permis de constituer une population totale de 2252 éléments qui, à cause des éliminations, est un peu plus petite que le maximum théorique de $2 \times 144 \times 8 = 2304$.

La perte de 52 EET est due à l'élimination de sept compteurs, marqués dans les tableaux par des astérisques. La population totale se compose donc de:

- a) Huit "tables de mai" (analogues à IA) contenant 1128 EET calculées pour les périodes de six mois: novembre à mai.

- b) Huit "tables de novembre", contenant 1124 EET calculées pour les périodes de six mois: mai à novembre.

Les erreurs relevées en mai proviennent de la période qui contient les mois d'hiver de sorte que cette période est appelée "saison froide" (en anglais "cold season") et est désignée par l'indice inférieur "c". La période de mai à novembre contient les mois d'été et est appelée "saison chaude" (en anglais "warm season") et est désignée par l'indice inférieur "w". Les valeurs individuelles des 2252 erreurs d'enregistrement totales ne sont pas citées dans le présent mémo; elles sont cependant donnée sous une forme classifiée dans la Fig 1 et Fig 2: dans la Fig 1 le total (2252) est considéré comme un "mélange" de deux sous-échantillons, partie "saison froide" et partie "saison chaude"; dans la Fig 2, le total est un mélange de la partie "direction froide" avec la partie "direction chaude". Les statistiques de la subdivision en "saisons" sont les suivantes:

		écart
Saison froide:	$N_c = 1128, \bar{m}_c = 0.57,$	type $s_c = 0.52.$
Saison chaude:	$N_w = 1124, \bar{m}_w = 0.22,$	" $s_w = 0.59.$
Total:	$N_t = 2252, \bar{m}_t = 0.39,$	" $s_t = 0.58.$

En ce qui concerne "direction", le nord et l'est (ensemble) constituent la direction froide, tandis que sud et ouest (ensemble) constituent la direction chaude:

Direction froide (N + E) :	$N_c = 1129, \bar{m}_c = 0.45,$	écart type $s_c = 0.54.$
Direction chaude (S + W) :	$N_w = 1123, \bar{m}_w = 0.33,$	" " $s_w = 0.62.$
Total:	$N_t = 2252, \bar{m}_t = 0.39,$	" " $s_t = 0.58.$

Avant de tirer les conclusions des Fig 1 et 2, il faut mentionner de nouveau qu'au cours des huit années d'expérience sept compteurs ont été éliminés pour défauts de fonctionnement. La manière dont les relevés ont été

accomplis ne permet pas de se rendre compte quand les compteurs éliminés ont commencé à mal fonctionner. Dans la moitié des cas, le compteur se trouve simplement et définitivement stoppé. Dans d'autres cas, le chiffre indiqué par le compteur n'avait aucun sens. Seulement un manufacturier expérimenté et un expert en mesures électriques peut évaluer la qualité des compteurs qui ont été testés de la manière qui est décrite dans le présent rapport. Si le nombre des cas où les compteurs ont mal fonctionné n'est pas considéré comme très significatif, on peut dire que les compteurs ont fonctionné d'une manière satisfaisante. Un compteur semble être, dans les limites de la tolérance admise, un instrument auquel on peut se fier malgré le dur climat de la région d'Ottawa.

Au sujet de la distribution des fréquences on peut faire la remarque générale suivante: les cinq diagrammes observés sont raisonnablement réguliers, symétriques et en forme de cloche mais quatre d'entre eux ne sont pas vraiment bien représentés par la courbe de Gauss. Ils sont leptokurtiques avec différents degrés de leptokurtose.* Dans la théorie moderne la présence de la leptokurtose dans les diagrammes des erreurs est traitée simplement comme un caractère additionnel sans influence spécifique sur la qualité des observations.

L'impression générale (suggérée par un examen superficiel) produite par la Fig 1 est celle de la "dispersion" tandis que celle produite par la Fig 2 est celle de la "cohésion". Ceci suggère l'idée que le critère "saison" agit sur le comportement des compteurs plus fortement que le critère "direction". Un examen plus profond montre que dans les deux figures les moyennes avec l'index w sont plus proches de 0 (zéro) que les moyennes avec l'indice c: il semble par conséquent que dans les conditions plus chaudes (sud et été) les erreurs d'enregistrement ont la tendance de décroître tandis que

* Note: un diagramme de fréquences est dit "leptokurtique" (terme de K. Pearson) si, dans sa portion centrale, les fréquences observées (représentées par les points . dans les deux totaux) sont plus grandes que les correspondantes fréquences indiquées par la courbe de Gauss (courbe qui représente le diagramme). Dans la courbe qui représente le total (2252) il y a neuf de ces points. Cet excès au centre est compensé par un abaissement correspondant des points situés des deux côtés de la partie centrale.

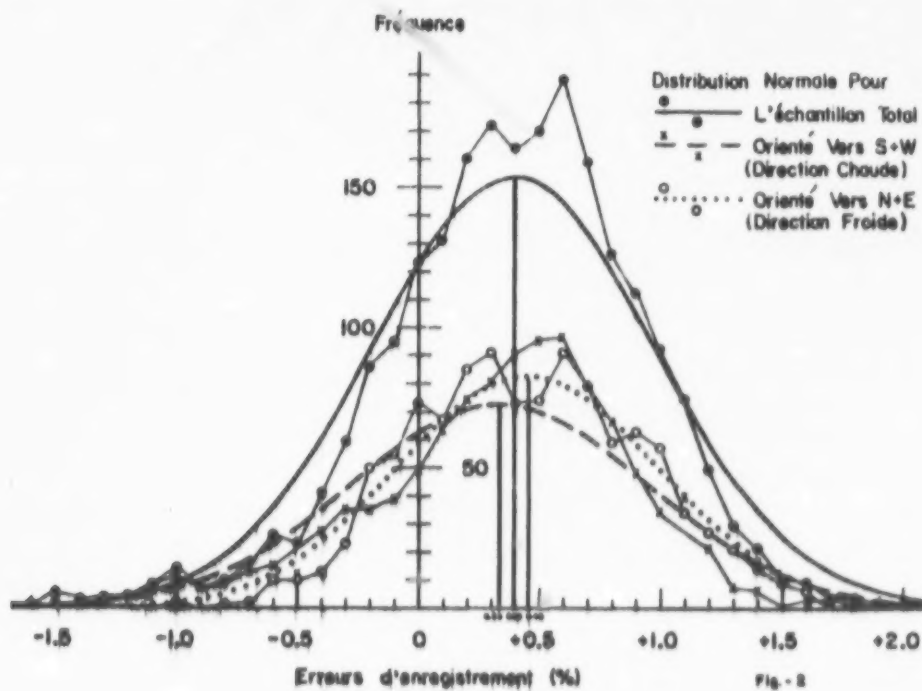
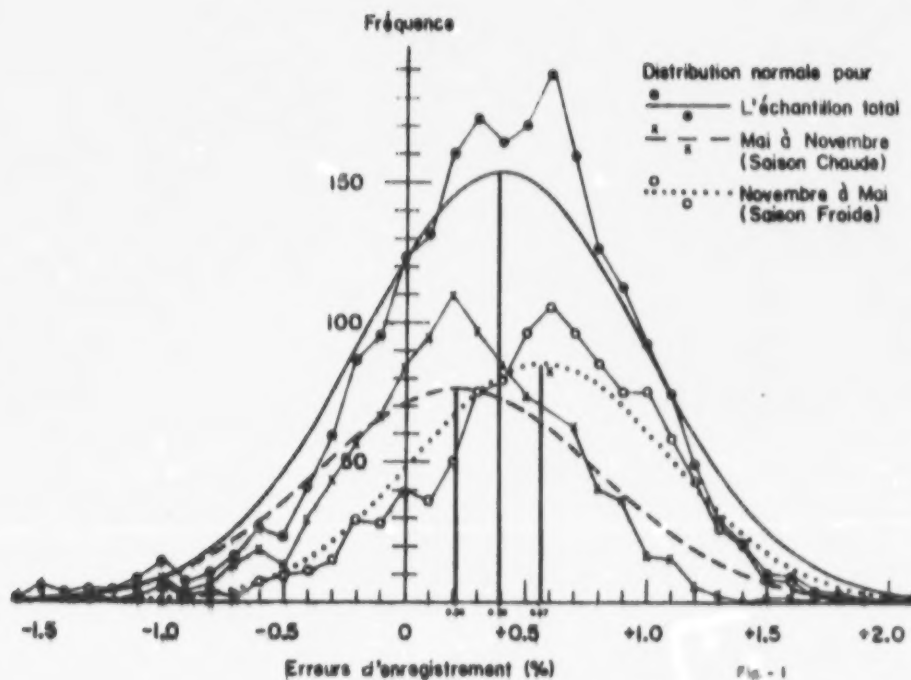
dans les conditions plus froides la tendance est dans le sens contraire. En plus, les moyennes des sous-échantillons sont plus éloignées l'une de l'autre dans la Fig 1 que dans la Fig 2. Et la différence dans Fig 1 est loin d'être non significative:

$$\overline{m}_c - \overline{m}_w = 0.35.$$

Comme s_c et s_w sont tous les deux de l'ordre de 0.55, la différence est donc de l'ordre de 0.6 écart type. Elle n'est donc pas négligeable et on peut dire: il est à l'avantage du consommateur de placer le compteur sur le côté sud du bâtiment et de consommer de l'électricité surtout pendant la période chaude. Cependant, son gain ne sera que "probable" et, de toute façon, seulement léger!

NOTE

L'échantillon d'Erreurs d'enregistrement totales (EET), analysé ci-dessus, est aussi utilisé pour illustrer la théorie des "Mélange", en particulier quand les composantes sont du type normal modulé. L'article sur ce sujet est publié dans le Bulletin Géodésique 53, No. 1, 1983, p. 62 et peut être obtenu des auteurs (MR ou EG) en s'adressant à NRC, Division de Physique, Ottawa, Ontario, Canada K1A 0R6.





Appendice I

Cet appendice contient les tables des données observées et les "sorties" (outputs) pour chacune des années 1972 - 1979. Dans ces tables les données observées sont présentées comme dans la Table IA.

Le lecteur remarquera que les résultats numériques, quoique un peu différents d'une année à l'autre, conduisent aux mêmes conclusions que celles présentées à la fin de la Section 2.

Comme les tables ont été faites par ordinateur, le texte n'est disponible qu'en anglais. Voici la liste des termes utilisés:

Date of Data: Date de relevé des indications
Criterion-Manufacturer: Critère-Manufacturier
Criterion-Direction: Critère-Direction
Square: Carré
East: Est
West: Ouest
North: Nord
South: Sud

TABLE I-8

Date of Data is 1972 NOV

	GE	FP	W	S
SOUTH	-.01 -.20 .12 .39 .07 -.17 .16 -.01 .54	.72 .56 -.95 .83 .24 -.40 .44 -.91 -.49	.62 .60 .72 .76 .14 -.72 .73 1.06 .87	1.24 1.48 .72 1.15 .80 .62 .64 1.02 .95
EAST	.31 .30 .11 .65 .36 .58 .56 .65 .47	-.13 .27 .43 .34 .07 -.07 .40 .53 .05	.05 -.29 -.21 -.38 1.17 .60 -.14 -.18 -.24	.78 .64 1.20 .57 .89 .75 .92 .87 .81
NORTH	.32 .11 .07 .65 .72 .69 .68 .62 .35	.26 .69 .98 .55 .88 .30 .62 .31 .42	.39 -.07 .68 -.18 .59 -.03 .98 .31 .03	.51 .60 .98 .44 .72 .25 .92 .81 1.20
WEST	.22 .46 .36 .05 .42 .99 .40 .40 .63	.50 .66 .30 .77 -.03 .50 .04 .20 .32	.27 .40 .03 .28 .15 -.43 .56 .57 -.08	.27 2.50 .59 .45 .47 .56 1.65 .88 .45
	t_1	t_2	t_3	t_4

Criterion- MANUFACTURER

N= 144.00
h=4

S= 58.23
M= 27.41

S
30.14
908.43

W
9.46
82.42

FP
10.20
104.84

Manufacturer
Totals (t)
Totals squared
Mu

N-h= 140.00
h-1=3

t zero = 62.82
t zero squ. = 3946.35
Mu zero = 35.32
A = 30.82
B = 22.91
C = 7.91
s' square = .16
S' square = 2.64
F = 16.12
(v' 3, v' = 140.00)

Criterion- DIRECTION

N= 144.00
h=4

S= 58.23
M= 27.41

WEST
16.76
280.90
7.80

NORTH
18.35
336.72
9.35

EAST
14.89
198.53
5.51

Direction
Totals (t)
Totals squared
Mu

N-h= 140.00
h-1=3

t zero = 62.82
t zero squ. = 3946.35
Mu zero = 27.82
A = 30.82
B = 30.40
C = .42
s' square = .22
S' square = .14
F = 1.56
(v' 3, v' = 140.00)

TABLE II

Date of Data is 1973 NOV		Date of Data is 1973 NOV		Criterion- MANUFACTURER	
				N= 144.00 h=4	
				S= 63.06 M= 33.64	
				Manufacturer GE	
				Totals (t) 14.06	
				Totals squared 197.68	
				Mu	
				FP	
				15.37	
				236.24	
				8.56	
				W	
				13.27	
				176.87	
				4.85	
				S	
				26.90	
				733.63	
				20.10	
				N-h= 140.00 h-1=3	
				t zero = 69.60	
				t zero squ. = 4844.16	
				Mu zero = 37.05	
				A = 29.42	
				B = 26.01	
				C = 3.41	
				s' square = .19	
				S' square = 1.14	
				F = 6.11	
				(v' 3, v' = 140.00)	
				N= 144.00 h=4	
				S= 63.06 M= 33.64	
				Direction SOUTH	
				Totals (t) 17.94	
				Totals squared 321.84	
				Mu	
				FAST	
				18.11	
				327.97	
				9.11	
				NORTH	
				20.37	
				414.94	
				11.53	
				WEST	
				13.18	
				173.71	
				4.83	
				N-h= 140.00 h-1=3	
				t zero = 69.60	
				t zero squ. = 4844.16	
				Mu zero = 34.40	
				A = 29.42	
				B = 28.66	
				C = .76	
				s' square = .20	
				S' square = .25	
				F = .81	
				(v' 3, v' = 140.00)	

TABLE III

Date of Data is 1974 NOV		Date of Data is 1974 NOV		Criterion- MANUFACTURER	
				N= 143.00 h=4	
				S= 47.27 M= 16.67	
				Manufacturer GE	
				Totals (r) 7.31	
				Totals squared 53.44	
				Mu 1.76	
				FP 6.99	
				48.86	
				1.36	
				W 12.22	
				149.33	
				4.15	
				S 22.30	
				497.29	
				13.81	
				N-h= 139.00 h-1=3	
				t zero = 48.82	
				t zero squar. = 2383.39	
				Mu zero = 20.85	
				A = 30.60	
				B = 26.42	
				C = 4.18	
				s' square = .19	
				S' square = 1.39	
				F = 7.33	
				(v' 3, v'' = 139.00)	
				N= 143.00 h=4	
				S= 47.27 M= 16.67	
				Direction SOUTH	
				Totals (r) 7.97	
				Totals squared 63.52	
				Mu 1.76	
				t zero = 48.82	
				t zero squar. = 2383.39	
				Mu zero = 17.66	
				A = 30.60	
				B = 29.61	
				C = .99	
				s' square = .21	
				S' square = .33	
				F = .64	
				(v' = 3 v'' = 139.00)	
				N-h= 139.00 h-1=3	
				FAST 13.13	
				172.40	
				4.79	
				NORTH 16.30	
				265.69	
				7.38	
				WEST 11.42	
				130.42	
				3.73	

TABLE IV

Date of Data is 1975 NOV		Date of Data is 1975 NOV		Criterion- MANUFACTURER	
-03	.76	.96	.70	S= 40.12	N= 141.00
-19	.21	.72	.62	M= 16.10	h=4
06	-16	.05	.59		
42	.37	.79	.81		
-07	.58	888	1.02	Manufacturer	GE
-42	-72	.34	.47	Totals (t)	6.94
08	.29	.63	.62	Totals squared	48.16
-05	-27	1.02	.84	Mu	1.38
.49	-.94	.62	.78		
				t zero	= 47.64
-51	.29	.35	1.16	t zero squar. =	2269.57
.24	.42	.11	.61	Mu zero	= 20.93
11	.50	.09	1.09	A	= 24.03
.55	.25	-14	.46	B	= 19.19
.31	-.01	1.55	.80	C	= 4.83
.37	.83	.65	.67	s' square	= .14
.49	.41	-.02	.96	s'' square	= 1.61
.53	.49	-11	.89	F	= 11.50
.44	-.14	.19	.81	(v' 3, v'' =	137.00)
.13	.05	.30	.39	S= 40.12	N= 141.00
-06	.74	-10	.31	M= 16.10	h=4
-08	.78	.44	.83		
.40	.50	-.01	.36	Direction	SOUTH
.49	.79	.67	.61	Totals (t)	11.99
.47	.13	.15	.09	Totals squared	143.76
.39	.59	-1.00	.81	Mu	4.11
.36	.14	.26	.90		
.21	.51	.24	1.16	t zero	= 47.64
				t zero squar. =	2269.57
.02	.28	.41	.17	Mu zero	= 17.06
.28	.37	.53	.53	A	= 24.03
.18	-.83	-11	.37	B	= 23.06
1888	.46	.26	.25	C	= .96
.29	-.29	.03	.3	s' square	= .17
.72	.05	-.26	.33	s'' square	= .3
.22	-.36	.44	.60	F	= 11.52
.10	-.15	.59	.66	(v' 3, v'' =	137.00)
.47	-.02	888	.51		
				FAST	NORTH
				15.71	12.95
				246.80	167.70
				6.86	4.66
					WEST
					6.99
					48.86
					1.44

TABLE V

Date of Data is 1976 NOV		Date of Data is 1976 NOV		Criterion- MANUFACTURER	
				N= 140.00 h=4	
.19	1.05	1.23	.85	Manufacturer	GE
-.05	.37	.74	.69	Totals (t)	12.05
.18	-.59	.23	.75	Totals squared	145.20
.56	.61	.86	.87		
-.01	.80	-.08	1.07		
-.21	-.09	.33	.61		
.32	.35	.61	.79		
.02	-1.16	1.05	.95		
.64	-.95	.61	.92		
				N-h= 136.00 h-1=3	
###	-.15	.67	1.18	t zero	67.40
.36	.55	.02	.70	t zero squar.	4542.76
.24	.68	.11	1.07	Mu zero	37.20
.69	.36	-.18	.52	A	25.79
.37	.11	1.50	.86	B	21.05
###	.91	.72	.78	C	4.75
.61	.45	-.05	1.10	s' square	.15
.66	.44	-.16	1.11	s' square	1.58
.57	-.16	.21	1.21	F	10.23
				(v' 3, v' 4)	136.00
				N= 140.00 h=4	
.29	.25	.46	.49	S= 58.24	
.09	.95	.05	.46	M= 32.45	
.03	.98	.67	1.01		
.53	.49	.05	.45		
.68	.93	.74	.68		
.66	.31	.26	.21		
.60	.72	.82	1.00		
.54	.21	.39	1.15		
.28	.60	.30	1.31		
				Criterion- DIRECTION	
.14	.46	.57	.37	SOUTH	
.45	.76	.78	.57	Direction	15.11
.27	1.33	.08	.62	Totals (t)	228.31
.01	.70	.29	.44	Totals squared	6.34
###	-.21	.36	.52	Mu	
.97	.33	-.23	.55	t zero	67.40
.36	-.30	.35	.86	t zero squar.	4542.76
.34	-.17	.60	.88	Mu zero	32.91
.67	.13	###	.82	A	25.79
				B	25.33
				C	.46
				s' square	.19
				s' square	.15
				F	1.21
				(v' 3, v' 4)	136.00
				N= 140.00 h=4	
				Criterion- MANUFACTURER	
				N= 140.00 h=4	
				Manufacturer	
				Totals (t)	
				Totals squared	
				FP	
				12.05	
				145.20	
				W	
				14.88	
				221.43	
				S	
				28.42	
				892.74	
				WFST	
				14.59	
				212.87	
				6.26	

TABLE VII

TABLE VII									
Date of Data is 1978 NOV									
Criterion- MANUFACTURER									
N= 137.00 h=4					M				
S= 32.69 M= 2.73					FP				
Manufacturer					GE				
Totals (t)					-1.12				
Totals squared					1.25				
Mu					.04				
t zero					19.34				
t zero squar.					374.04				
Mu zero					9.68				
A					29.9				
B					23.01				
C					6.95				
S' square					.17				
S'' square					2.32				
F					13.38				
(v' 3, v''=					133.00)				
N-h= 133.00 h-1=3									
S= 32.69 M= 2.73					N= 137.00 h=4				
Direction					SOUTH				
Totals (t)					-19				
Totals squared					.04				
Mu					.00				
t zero					19.34				
t zero squar.					374.04				
Mu zero					4.00				
A					29.96				
B					28.61				
C					1.35				
S' square					.22				
S'' square					.45				
F					.48				
(v' 3, v''=					133.00)				
N-h= 133.00 h-1=3					N-h= 133.00 h-1=3				
S= 32.69 M= 2.73					S= 32.69 M= 2.73				
Direction					EAST				
Totals (t)					6.19				
Totals squared					38.32				
Mu					1.13				
t zero					19.34				
t zero squar.					374.04				
Mu zero					4.00				
A					29.96				
B					28.61				
C					1.35				
S' square					.22				
S'' square					.45				
F					.48				
(v' 3, v''=					133.00)				
N-h= 133.00 h-1=3					N-h= 133.00 h-1=3				
S= 32.69 M= 2.73					S= 32.69 M= 2.73				
Direction					NORTH				
Totals (t)					9.04				
Totals squared					81.72				
Mu					2.40				
t zero					19.34				
t zero squar.					374.04				
Mu zero					4.00				
A					29.96				
B					28.61				
C					1.35				
S' square					.22				
S'' square					.45				
F					.48				
(v' 3, v''=					133.00)				
N-h= 133.00 h-1=3					N-h= 133.00 h-1=3				
S= 32.69 M= 2.73					S= 32.69 M= 2.73				
Direction					WEST				
Totals (t)					4.30				
Totals squared					18.49				
Mu					.54				
t zero					19.34				
t zero squar.					374.04				
Mu zero					4.00				
A					29.96				
B					28.61				
C					1.35				
S' square					.22				
S'' square					.45				
F					.48				
(v' 3, v''=					133.00)				
N-h= 133.00 h-1=3					N-h= 133.00 h-1=3				
S= 32.69 M= 2.73					S= 32.69 M= 2.73				
Direction					WEST				
Totals (t)					4.30				
Totals squared					18.49				
Mu					.54				
t zero					19.34				
t zero squar.					374.04				
Mu zero					4.00				
A					29.96				
B					28.61				
C					1.35				
S' square					.22				
S'' square					.45				
F					.48				
(v' 3, v''=					133.00)				
N-h= 133.00 h-1=3					N-h= 133.00 h-1=3				
S= 32.69 M= 2.73					S= 32.69 M= 2.73				
Direction					WEST				
Totals (t)					4.30				
Totals squared					18.49				
Mu					.54				
t zero					19.34				
t zero squar.					374.04				
Mu zero					4.00				
A					29.96				
B					28.61				
C					1.35				
S' square					.22				
S'' square					.45				
F					.48				
(v' 3, v''=					133.00)				
N-h= 133.00 h-1=3					N-h= 133.00 h-1=3				
S= 32.69 M= 2.73					S= 32.69 M= 2.73				
Direction					WEST				
Totals (t)					4.30				
Totals squared					18.49				
Mu					.54				
t zero					19.34				
t zero squar.					374.04				
Mu zero					4.00				
A					29.96				
B					28.61				
C					1.35				
S' square					.22				
S'' square					.45				
F					.48				
(v' 3, v''=					133.00)				
N-h= 133.00 h-1=3					N-h= 133.00 h-1=3				
S= 32.69 M= 2.73					S= 32.69 M= 2.73				
Direction					WEST				
Totals (t)					4.30				
Totals squared					18.49				
Mu					.54				
t zero					19.34				
t zero squar.					374.04				
Mu zero					4.00				
A					29.96				
B					28.61				
C					1.35				
S' square					.22				
S'' square					.45				
F					.48				
(v' 3, v''=					133.00)				
N-h= 133.00 h-1=3					N-h= 133.00 h-1=3				
S= 32.69 M= 2.73					S= 32.69 M= 2.73				
Direction					WEST				
Totals (t)					4.30				
Totals squared					18.49				
Mu					.54				
t zero					19.34				
t zero squar.					374.04				
Mu zero					4.00				
A					29.96				
B					28.61				
C					1.35				
S' square					.22				
S'' square					.45				
F					.48				
(v' 3, v''=					133.00)				
N-h= 133.00 h-1=3					N-h= 133.00 h-1=3				
S= 32.69 M= 2.73					S= 32.69 M= 2.73				
Direction					WEST				
Totals (t)					4.30				
Totals squared					18.49				
Mu					.54				
t zero					19.34				
t zero squar.					374.04				
Mu zero					4.00				
A					29.96				
B					28.61				
C					1.35				
S' square					.22				
S'' square					.45				
F					.48				
(v' 3, v''=					133.00)				
N-h= 133.00 h-1=3					N-h= 133.00 h-1=3				
S= 32.69 M= 2.73					S= 32.69 M= 2.73				
Direction					WEST				
Totals (t)					4.30				
Totals squared					18.49				
Mu					.54				
t zero					19.34				
t zero squar.					374.04				
Mu zero					4.00				
A					29.96				
B					28.61				
C					1.35				
S' square					.22				
S'' square					.45				
F					.48				
(v' 3, v''=					133.00)				
N-h= 133.00 h-1=3					N-h= 133.00 h-1=3				
S= 32.69 M= 2.73					S= 32.69 M= 2.73				
Direction					WEST				
Totals (t)					4.30				
Totals squared					18.49				
Mu					.54				
t zero					19.34				
t zero squar.					374.04				
Mu zero					4.00				
A					29.96				
B					28.61				
C					1.35				
S' square					.22				
S'' square					.45				
F					.48				
(v' 3, v''=					133.00)				
N-h= 133.00 h-1=3					N-h= 133.00 h-1=3				
S= 32.69 M= 2.73					S= 32.69 M= 2.73				
Direction					WEST				
Totals (t)					4.30				
Totals squared					18.49				
Mu					.54				
t zero					19.34				
t zero squar.					374.04				
Mu zero					4.00				
A					29.96				
B					28.61				
C					1.35				
S' square					.22				
S'' square					.45				
F					.48				
(v' 3, v''=					133.00)				
N-h= 133.00 h-1=3					N-h= 133.00 h-1=3				
S= 32.69 M= 2.73					S= 32.69 M= 2.73				
Direction					WEST				
Totals (t)					4.30				
Totals squared					18.49				
Mu					.54				
t zero					19.34				
t zero squar.					374.04				
Mu zero					4.00				
A					29.96				
B					28.61				
C					1.35				
S' square					.22				
S'' square					.45				
F					.48				
(v' 3, v''=					133.00)				
N-h= 133.00 h-1=3					N-h= 133.00 h-1=3				
S= 32.69 M= 2.73					S= 32.69 M= 2.73				
Direction					WEST				
Totals (t)					4.30				
Totals squared					18.49				
Mu					.54				
t zero					19.34				
t zero squar.					374.04				
Mu zero					4.00				
A					29.96				
B					28.61				
C					1.35				
S' square					.22				
S'' square					.45				
F					.48				
(v' 3, v''=					133.00)				
N-h= 133.00 h-1=3					N-h= 133.00 h-1=3				
S= 32.69 M= 2.73					S= 32.69 M= 2.73				
Direction					WEST				
Totals (t)					4.30				
Totals squared					18.49				
Mu					.54				
t zero					19.34				
t zero squar.					374.04				
Mu zero					4.00				
A					29.96				
B					28.61				
C					1.35				
S' square					.22				
S'' square					.45				
F					.48				
(v' 3, v''=					133.00)				
N-h= 133.00 h-1=3					N-h= 133.00 h-1=3				
S= 32.69 M= 2.73					S= 32.69 M= 2.73				
Direction					WEST				
Totals (t)					4.30				
Totals squared					18.49				
Mu					.54				
t zero					19.34				
t zero squar.					374.04				
Mu zero					4.00				
A					29.96				
B					28.61				
C					1.35				
S' square					.22				
S'' square					.45				
F					.48				
(v' 3, v''=					133.00)				
N-h= 133.00 h-1=3					N-h= 133.00 h-1=3				
S= 32.69 M= 2.73					S= 32.69 M= 2.73				
Direction					WEST				
Totals (t)					4.30				
Totals squared					18.49				
Mu					.54				
t zero					19.34				
t zero squar.					374.04				
Mu zero					4.00				
A					29.96				
B					28.61				
C					1.35				
S' square					.22				
S'' square					.45				
F					.48				
(v' 3, v''=					133.00)				
N-h= 133.00 h-1=3					N-h= 133.00 h-1=3				
S= 32.69 M= 2.73					S= 32.69 M= 2.73				
Direction					WEST				
Totals (t)					4.30				
Totals squared					18.49				
Mu					.54				
t zero					19.34				
t zero squar.					374.04				
Mu zero					4.00				
A					29.96				
B					28.61				
C					1.35				
S' square					.22				
S'' square					.45				
F					.48				
(v' 3, v''=					133.00)				
N-h= 133.00 h-1=3					N-h= 133.00 h-1=3				
S= 32.69 M= 2.73					S= 32.69 M= 2.73				
Direction					WEST				
Totals (t)					4.30				
Totals squared					18.49				
Mu					.54				
t zero					19.34				
t zero squar.					374.04				
Mu zero					4.00				
A					29.96				
B					28.61				
C					1.35				
S' square					.22				
S'' square					.45				
F					.48				
(v' 3, v''=					133.00)				
N-h= 133.00 h-1=3					N-h= 133.00 h-1=3				
S= 32.69 M= 2.73					S= 32.69 M= 2.73				
Direction					WEST				
Totals (t)					4.30				
Totals squared					18.49				
Mu					.54				
t zero					19.34				
t zero squar.					374.04				
Mu zero					4.00				
A					29.96				
B					28.61				
C					1.35				
S' square					.22				
S'' square					.45				
F					.48				
(v' 3, v''=					133.00)				
N-h= 133.00 h-1=3					N-h= 133.00 h-1=3				
S= 32.69 M= 2.73					S= 32.69 M= 2.73				
Direction					WEST				
Totals (t)					4.30				
Totals squared					18.49				
Mu					.54				
t zero					19.34				
t zero squar.					374.04				
Mu zero					4.00				
A					29.96				
B					28.61				
C					1.35				
S' square					.22				
S'' square					.45				
F					.48				
(v' 3, v''=					133.00)				
N-h= 133.00 h-1=3					N-h= 133.00 h-1=3				
S= 32.69 M= 2.73					S= 32.69 M= 2.73				
Direction					WEST				
Totals (t)					4.30				
Totals squared					18.49				
Mu					.54				
t zero					19.34				
t zero squar.					374.04				
Mu zero					4.00				
A					29.96				
B					28.61				
C					1.35				
S' square					.22				
S'' square					.45				
F					.48				
(v' 3, v''=					133.00)				
N-h= 133.00 h-1=3					N-h= 133.00 h-1=3				
S= 32.69 M= 2.73					S= 32.69 M= 2.73				
Direction					WEST				
Totals (t)					4.30				
Totals squared					18.49				
Mu					.54				
t zero					19.34				
t zero squar.					374.04				
Mu zero					4.00				
A					29.96				
B					28.61				
C					1.35				
S' square					.22				
S'' square					.45				
F					.48				
(v' 3, v''=					133.00)				
N-h= 133.00 h-1=3					N-h= 133.00 h-1=3				
S= 32.69 M= 2.73					S= 32.69 M= 2.73				
Direction					WEST				
Totals (t)					4.30				
Totals squared					18.49				
Mu					.54				
t zero					19.34				
t zero squar.					374.04				
Mu zero					4.00				
A					29.96				
B					28.61				
C					1.35				

TABLE VIII

Date of Data is 1979 NOV		Date of Data is 1979 NOV		TABLE VIII	
		Criterion- MANUFACTURER			
.09 - .14 *** .54 - .11 *** .18 - .02 .63	1.16 .21 - .69 .45 - .83 - .38 - 1.46 - 1.15	S = 63.47 M = 31.70	71 .56 .70 .84 .27 .79 .8 .85	M = 138.00 h = 4	FP 10.06 161.20 2.81 18.13 328.70 9.39 28.09 789.05 21.92
Manufacturer					
Totals (t)					
Totals squared					
Mu					
t zero					
t zero squar. =					
Mu zero					
A					
B					
C					
s' square					
S' square					
(v' 3, v' 4)					
M-h = 134.00 h-1 = 3					
Criterion- DIRECTION					



Appendice II

Note sur la corrélation entre l'erreur d'enregistrement totale (EET) et l'erreur du disque dans le projet "Compteurs sur le toit".

La présente Note a été préparée en coopération avec B. Fulsom, chef du Laboratoire d'Electricité et J. Samuel, étudiant, engagé temporaire pendant l'été 1982.



LES ERREURS DU DISQUE (ED) DANS
LE PROJET "COMPTEURS SUR LE TOIT"

1. Introduction

La principale caractéristique des "erreurs d'enregistrement totales" (EET) est qu'elles demandent que les mesures consécutives soient séparées l'une de l'autre par des intervalles de temps considérables, au moins de plusieurs mois. Ceci entraîne la consommation de grandes quantités d'énergie électrique et le maintien, pendant de longues périodes, de la circuiterie et des installations appropriées. L'une des raisons pour cet état de choses est qu'un compteur ne fournit une lecture bien définie que s'il indique une quantité relativement considérable de watt-heures. Il va de soi que pour le manufacturier et pour l'utilisateur qui consomme une grande quantité d'énergie électrique, les EET sont d'une importance capitale et doivent être surveillées.

Dans la pratique, cependant, l'une des opérations courantes est la détermination de l'erreur d'enregistrement sur une période très courte, p.ex. de l'ordre de quelques minutes. Il est évident que, dans ce cas, une détermination ne peut pas être aussi sûre que celle de l'EET mais, pour des raisons évidentes, elle doit aussi pouvoir être effectuée. Lorsqu'un usager trouve que son compteur demande à être vérifié, ce compteur est acheminé vers un laboratoire dans lequel l'erreur du disque (ED) est aussitôt déterminée.

La propriété fondamentale d'un compteur à induction est qu'un tour de son disque tournant correspond à un nombre constant d'unités d'énergie (joule ou watt-heure). Cette constante, désignée par le symbole " K_h ", est un peu

perturbée par les conditions dans lesquelles la mesure est effectuée. La détermination de EET bénéficie de l'effet égalisateur dû au grand nombre de tous qui ont lieu pendant de long intervalles de temps.

Un autre point important qui doit être mentionné concerne le courant utilisé pendant les déterminations du ED. L'influence de son intensité sur K_h n'est pas simple, et en tout cas pas linéaire. Dans la pratique courante, K_h est déterminé avec des courants égaux à 2.5 A et 25A. Tous les manufacturiers s'efforcent à produire des compteurs qui possèdent des valeurs prescrites pour les constantes du disque et un compteur est considéré (par le manufacturier) comme acceptable dans les installations domestiques si son erreur du disque (ED) n'est pas supérieure à 1%. En fait, cette limite est plus étroite que la tolérance légale (2%) car la vérification des nouveaux compteurs est basée sur la méthode "par échantillonnage" et non sur la vérification de chaque instrument individuel.

La référence pour les ED est un instrument appelé "Étalon de transfert". Il produit des battements, le temps entre deux battements consécutifs correspondant à une quantité constante d'énergie électrique. Il est utilisé dans le laboratoire de Métrologie légale pour déterminer la constante K_h du "Compteur de Référence". La constante K_h est ajustée de façon qu'un watt-heure corresponde à trois révolutions du disque. En d'autres termes, $K_h = 1/3$ et, légalement, un compteur est acceptable si son erreur ne dépasse pas 2%.

2. Travail expérimental et calcul des corrélations

Il est évident que même sur un compteur de haute qualité la relation linéaire entre ses EET et ED ne sera pas absolument parfaite. Dans le langage statistique nous pouvons tout de même prévoir qu'entre ces deux sortes d'erreur la corrélation sera, sinon "parfaite", en tout cas très étroite. Au

cours de l'expérience décrite ci-dessus, qui a duré huit années consécutives, un certain nombre de déterminations des erreurs du disque ont été effectuées. Toutefois, pour des raisons pratiques, l'ensemble des ED ainsi obtenues n'a pas pu être traité dans sa totalité; seulement trois périodes ont pu être calculées et statistiquement analysées. La première et la troisième période s'étendent sur une année chacune; la seconde période occupe deux années consécutives. L'erreur du disque pour une période est la moyenne des observations effectuées au commencement, au milieu et à la fin de la période. Cette moyenne est désignée par le symbole $\bar{X}_i$ de sorte que dans chaque période, à une valeur Y_i correspond une valeur $\bar{X}_i$. Les tables des valeurs individuelles X_i ne sont pas citées dans le présent memorandum et les données concernant les périodes sont résumées comme suit.

Première période 1 Nov 1971 à 11 Oct 1972, N = 144.
Deuxième " 15 Oct 1974 à 15 Oct 1976, N = 139.
Troisième " 12 Oct 1975 à 15 Oct 1979, N = 138.

La suite des opérations conduisant à la valeur de l'index de corrélation r est:

A) Calcul des moyennes $\bar{Y}$ and $\bar{X}$:

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i ; \bar{X} = \frac{1}{N} \sum_{i=1}^N \bar{X}_i$$

B) Calcul des déviations y_i et x_i :

$$y_i = Y_i - \bar{Y} ; x_i = \bar{X}_i - \bar{X}.$$

C) Calcul des variances s_y^2 et s_x^2 :

$$s_y^2 = \frac{1}{N} \sum_{i=1}^N y_i^2 ; s_x^2 = \frac{1}{N} \sum_{i=1}^N x_i^2.$$

D) Calcul de la covariance:

$$\mu_{11} = \frac{1}{N} \sum_{i=1}^N y_i x_i.$$

E) L'index de corrélation r , qui résulte des valeurs ci dessus est:

$$\mu_r = \frac{11}{s_y s_x}$$

Pour se rendre compte à quel point ce " r " est significatif il faut se rapporter à la valeur correspondante du variate t de Student. La relation entre t et r est:

$$t = \frac{r}{1-r^2} (N - 2)$$

Une forte corrélation (à laquelle on doit s'attendre) doit conduire à des valeurs de r qui sont proches de l'unité; ceci, à son tour, conduit aux grandes valeurs de t . Par exemple, pour $r = 0.5$ ($N = 144$), $t = 6.9$. Ce t est donc très grand et cela indique que la corrélation est très forte.

Les valeurs numériques des diverses quantités mentionnées ci dessus sont présentées dans la table suivante:

PREMIÈRE PÉRIODE (1972)	DEUXIÈME PÉRIODE (1976)	TROISIÈME PÉRIODE
$\mu_{11} = 0.170$	0.180	0.250
$s_y^2 = 0.215$	0.185	0.236
$s_y = 0.464$	0.430	0.486
$s_x^2 = 0.301$	0.324	0.361
$s_x = 0.549$	0.569	0.601
$r = 0.667$	0.736	0.855

Il faut remarquer que toutes les valeurs de r sont supérieures à 0.5. Pour $r = 0.667$, t est plus grand que 10. Les corrélations entre X et Y dans les trois périodes sont donc très fortes. La conclusion générale de

l'expérience et de l'analyse statistique est qu'il n'y a pas de raison d'invalider l'usage traditionnel de l'erreur du disque pour vérifier l'exactitude d'un compteur électrique.

Si les circonstances le permettent (c.à.d. si l'on dispose d'un personnel approprié) il sera peut-être possible d'appliquer l'analyse de la variance directement aux erreurs du disque. Il serait intéressant de voir comment cette analyse s'accorde, dans ses conclusions, avec les résultats de l'analyse de la variance appliquée aux erreurs d'enregistrement totales (EET).